数理解析研究所講究録
1503巻 2006年 160-163

New Pumping Lemma for Non-Linear Context-Free Languages[1]

GÉZA HORVÁTH
University of Debrecen,
Kyoto Sangyo University
e-mail: geza@inf.unideb.hu

Abstract

Pumping lemmas are created to prove that given languages are not belong to certain language classes. There are many pumping lemmas for context-free languages, in this paper we show a new pumping lemma, which is sometimes more effective than the classical Bar-Hilell lemma, and it has an interesting new property, it can be used to prove that a given language belongs to the class of linear languages.

1 Introduction

Pumping lemmas give necessary condition for a language to belong one of the language classes. In many case it is very simple to show that a language does not satisfy the condition of a pumping lemma, which means that the given language is not in the given language class. It is the classical application of the pumping lemmas.

The first pumping lemma is introduced by Bar-Hillel, Perles, and Shamir in 1961 for context-free languages. [1] Since that time many pumping lemmas are introduced for different language classes. Some of them are simple, some of them are more complicated. Sometimes a new pumping lemma is introduced to prove that a special language does not belong to a given language class, but mainly researchers try to create stronger and stronger pumping lemmas, which means that the new pumping lemma can be applied for more language than the previous pumping lemmas.

In this paper we introduce a new pumping lemma, which is different from the previous pumping lemmas, because it gives necessary condition for languages belonging to the non-linear context-free language class. We will show that we can not compare this lemma with the Bar-Hillel lemma, because this new pumping lemma sometimes stronger, and sometimes weaker then the classical pumping lemma for context-free languages. Finally, we show that how can we use this new lemma to prove that a given language belongs to the class of linear languages, which is a new application of pumping lemmas.

[1]This work was supported by grants of Japanese Society for Promotion of Science (P04028), and the Hungarian National Foundation for Scientific Research (OTKA T049409).

2 Preliminaries

A *word* is a finite sequence of elements of some finite nonempty set Σ. We call the set Σ an *alphabet,* the elements of Σ *letters.* The set of all words over Σ is denoted by Σ^*. We put $\Sigma^+ = \Sigma^* \setminus \{\lambda\}$, where λ denotes the *empty word* having no letters. If u and v are words over an alphabet Σ, then their *catenation* uv is also a word over Σ. Especially, for any word uvw, we say that v is a *subword* of uvw.

A *language* over Σ is a set $L \subseteq \Sigma^*$. We extend the concept of catenation for the class of languages as usual. Therefore, if L_1 and L_2 are languages, then their product is $L_1L_2 = \{p_1p_2 \mid p_1 \in L_1, p_2 \in L_2\}$. Let p be a word. We put $p^0 = \lambda$ and $p^n = p^{n-1}p$ $(n > 0)$. Thus p^k $(k \geq 0)$ is the *k-th power* of p. If there is no danger of confusion, then sometimes we identify p with the singleton set $\{p\}$. Thus we will write p^* and p^+ instead of $\{p\}^*$ and $\{p\}^+$, respectively.

An *(unrestricted generative, or simply, unrestricted) grammar* is an ordered quadruple $G = (N, \Sigma, S, P)$ where N and Σ are disjoint alphabets, $S \in \Sigma$, and P is a finite set of ordered pairs (U, V) such that V is a word over the alphabet $N \cup \Sigma$ and U is a word over $N \cup \Sigma$ containing at least one letter of N. The elements of N are called *variables* or *nonterminals,* and those of Σ *terminals.* $N \cup \Sigma$ is the *total alphabet* and S is called the *start symbol.* Elements (U, V) of P are called *productions* and are written $U \to V$. If $U \to V \in P$ implies $U \in N$ then G is called *context-free.* If $U \to V \in P$ implies $U \in N$ and $V \in \Sigma^*N\Sigma^*$ then G is called *linear.*

A word W over $N \cup \Sigma$ *derives directly* a word W', in symbols, $W \xrightarrow{1} W'$, if and only if there are words W_1, U, W_2, V such that $W = W_1UW_2, W' = W_1VW_2$ and $U \to V$ belongs to P. W *derives* W', or in symbols, $W \xrightarrow{*} W'$ if and only if there is a finite sequence of words $W_0, \ldots, W_k (k \geq 0)$ over $N \cup \Sigma$ with $W_0 = W, W_k = W'$ and $W_i \xrightarrow{1} W_{i+1}$ for $0 \leq i \leq k-1$. Thus for every $W \in (N \cup \Sigma)^*$ we have $W \xrightarrow{*} W$.

The set $S(G) = \{W \mid W \in (N \cup \Sigma)^*, S \xrightarrow{*} W\}$ is called the set of *sentential forms* of G. The *language* $L(G)$ *generated by* G is defined by $L(G) = S(G) \cap \Sigma^*$. $L \subseteq \Sigma^*$ is a *context-free language* if we have $L = L(G)$ for some context-free grammar G, and $L \subseteq \Sigma^*$ is a *linear language* if we have $L = L(G')$ for some linear grammar G'.

Lemma 1 (Bar-Hillel Lemma) [1] *Given a context-free language L there exists two integers n, m such that any word $w \in L$ where $|w| \geq n$, admits a factorization $w = uvwxy$ satisfying*

1. $uv^iwx^iy \in L$ *for all integer* $i \geq 0$
2. $|vx| \neq 0$
3. $|vwx| \leq m$.

Example: Let $L = \{a^ib^ic^i \mid i \geq 0\}$. With the Bar-Hillel Lemma it is easy to show that the language L is not context-free.

Lemma 2 (Pumping Lemma for Linear Languages) [2] *Given a linear language L there exists an integer n such that any word $w \in L$ where $|w| \geq n$, admits a factorization $w = uvwxy$ satisfying*

1. $uv^iwx^iy \in L$ *for all integer* $i \geq 0$
2. $|uvxy| \neq 0$
3. $|vwx| \leq n$.

Example: With the above lemma it is easy to show that the language $L = \{a^ib^ic^jd^j \mid i, j \geq 0\}$ is not linear.

3 Main Result

Theorem 1 (New Pumping Lemma) *Given a non-linear context-free language L there exists (infinite many) word $w \in L$ which admits a factorization $w = rstuvwxyz$ satisfying*

1. $rs^itu^ivw^jxy^jz \in L$ *for all integer* $i, j \geq 0$
2. $|su| \neq 1$
3. $|wy| \neq 1$.

Proof: Let $G = (V_N, V_T, S, P)$ context-free grammar such that $L(G) = L$, and let $G_A = (V_N, V_T, A, P)$ for all $A \in V_N$. Because L is non-linear, there exists $A, B \in V_N$ and $\alpha, \beta, \gamma \in V_T^*$ such that $S \xrightarrow{*} \alpha A \beta B \gamma$, where both of the languages $L(G_A)$ and $L(G_B)$ are infinite. Then the words $\alpha L(G_A) \beta L(G_B) \gamma \in L$, and applying the Bar-Hillel Lemma for $L(G_A)$ and $L(G_B)$ we receive that $\alpha u_1 v_1^i w_1 x_1^i y_1 \beta u_2 v_2^j w_2 x_2^j y_2 \gamma \in L$. Let $r = \alpha u_1$, $s = v_1$, $t = w_1$, $u = x_1$, $v = y_1 \beta u_2$, $w = v_2$, $x = w_2$, $y = x_2$, $z = y_2 \gamma$, and we have the above form. □

4 Applications

4.1 The classical application

Let

$$H \subseteq \{1^2, 2^2, 3^2, ...\}$$

infinite set, and let

$$L_H = \{a^kb^ka^lb^l\} \mid k, l \geq 1;\ k \in H \text{ or } l \in H\} \cup \{a^mb^m \mid m \geq 1\}.$$

The language L_H satisfies the Bar-Hillel condition, so we can not apply the Bar-Hillel Lemma to show that L_H is not context-free. However the L_H language does not satisfy the condition of the pumping lemma for linear languages, so L_H is not linear. At this point we can apply the new pumping lemma, and the language L_H does not satisfy the condition of the new pumping lemma. This means L_H is not context-free.

In this example the new pumping lemma was stronger than the Bar-Hillel Lemma. However we can show an example, where the Bar-Hillel Lemma is stronger than the new pumping lemma.

Let

$$L_B = \{a^i b^i c^j d^j \mid i, j \geq 0\} \cup \{d^k e^k f^k \mid k \geq 0\}.$$

The language L_B satisfies the condition of the pumping lemma for non-linear context-free languages, and does not satisfy the condition of the Bar-Hillel Lemma, so L_B also not context-free.

This means there are cases when the Bar-Hillel Lemma works, and the new lemma does not work, and there are cases when the new lemma works, and the Bar-Hillel Lemma does not work, so we can not say that one of them is stronger than the other, we have to try both.

4.2 The non-classical application

This new pumping lemma has an interesting, new application. A context-free language which does not satisfy the condition of the new pumping lemma is linear. This means if we would like to prove that a given language is linear, we can apply the new pumping lemma on the language, and if the language does not satisfy the condition of the new pumping lemma, and we can give a context-free grammar generating the language, that is a proof of the language is linear. See the following example.

Let

$$L = \{a^i b^i b^i \mid i \leq 0\}.$$

The language L does not satisfy the condition of the pumping lemma for non-linear context-free languages, and the $G = (\{S, B\}, \{a, b\}, S, \{S \to aSB, S \to \lambda, B \to bb\})$ context-free grammar generates L. This means the language L is linear.

5 Acknowledgement

The author is grateful to Sándor Horváth and Masami Ito for their helpful comments.

References

[1] Bar-Hillel, Y., Perles, M., Shamir, E., On formal properties of simple phrase structure grammars, *Z. Phonetik. Sprachwiss. Komm.*, **14**, (1961), 143-172.

[2] [2] Rozenberg, G., Salomaa, A., (Eds.) Handbook of Formal Languages, *Springer* (1997).