Theoretical Computer Science 255 (2001) 601–605

Theoretical Computer Science

www.elsevier.com/locate/tcs

Note

The growth function of context-free languages

Roberto Incitti[a,b,*]

[a] Dipartimento di Matematica, Università di Bologna, Piazza di Porta San Donato 5, 40127 Bologna, Italy

[b] Institut Gaspard Monge, Cité Descartes 77454 Marne-la-Vallée, Cedex, France

Received 29 June 1999; revised 23 December 1999; accepted 28 February 2000

Abstract

In this paper we show that the growth of a context-free language is either polynomial or exponential.

1. Introduction

Let L be a formal language on the finite alphabet Γ. For $w \in \Gamma^*$, denote by $|w|$ the number of letters of w. Set $\gamma_L(n) = \#\{x \in X,\ |x| \leqslant n\}$. The function $\gamma_L(n)$ is called the *growth function* of L. A classical result of Chomsky and Schützenberger [3] states that if L is context-free and unambiguous, then the series $\gamma_L(z) = \sum_{n>0} \gamma_L(n)z^n$ is algebraic. Moreover, if $\gamma_L(z)$ is algebraic, then the growth function of L is either polynomial or exponential, (in this case one says that the language is, respectively, of *polynomial growth* and of *exponential growth*). Flajolet [4] showed that there exist context-free languages for which the series $\sum_{n>0} \gamma_L(n)z^n$ is trascendental and raised the question as to whether there exist context-free languages of *intermediate growth*, that is, greater that any polynomial function and smaller that any exponential one. In [6] Grigorchuk and Machì gave an example of language of intermediate growth, recognizable by a one-way deterministic non-erasing stack automaton, which is, in a sense, very close to a context-free language. In this paper we show that a context-free language has a growth function that is either polynomial or exponential. The next step in this study

* Correspondence address: Dipartimento di Matematica, Università di Bologna, Piazza di Porta San Donato 5, 40127 Bologna, Italy.

E-mail address: incitti@dm.unibo.it (R. Incitti).

0304-3975/01/$ - see front matter
PII: S0304-3975(00)00152-3

could be to examine the frontier between the context-free languages and the class of languages that has been studied by Grigorchuk and Machì.

2. Grammars of exponential growth

In what follows, G will denote a grammar on the finite alphabet Γ, and whose set of nonterminal symbols is $\Sigma = \{S = S_1, \dots, S_n\}$. For $\alpha \in (\Sigma \cup \Gamma)^*$, we will write $S \stackrel{*}{\Rightarrow} \alpha$ if there exists a derivation $S \Rightarrow w_1 \cdots \Rightarrow \alpha$. We will also suppose G proper, that is, $n = 1$, or, for every $k \in [2, n]$, there exist $a_k, b_k \in \Gamma^*$ and a derivation $S \stackrel{*}{\Rightarrow} a_k S_k b_k$. We will denote by $L(G) \subseteq \Gamma^*$ the context-free language generated by the proper grammar G. If $a, b \in \Gamma^*$, we will denote by $\langle a, b\rangle$ the submonoid of Γ^* generated by a and b, and, for $w \in \langle a, b\rangle$, we will denote by $|w|_{a,b}$ the minimum length of a representation of w as a product of elements of $\{a, b\}$.

Definition 2.1. Let D, E be, respectively, the languages of the words x (respectively, y) $\in \Gamma^*$ for which there exists e_x (respectively d_y) $\in \Gamma^*$ such that $S \stackrel{*}{\Rightarrow} xSe_x$ (respectively, $S \stackrel{*}{\Rightarrow} d_ySy$).

Lemma 2.2. *Let G be a grammar. Let $d_1, d_2 \in D$ (respectively, $e_1, e_2 \in E$). Then, there exist $a \in \mathbb{N}$, $m \in L(G)$ such that, for every $w \in \langle d_1, d_2\rangle$, $|w|_{d_1,d_2} \leqslant n$, (respectively $w \in \langle e_1, e_2\rangle$, $|w|_{e_1,e_2} \leqslant n$), there exists $f_w \in \Gamma^*$ (respectively $d_w \in \Gamma^*$), such that $S \stackrel{*}{\Rightarrow} wmf_w$ (respectively, $S \stackrel{*}{\Rightarrow} d_wmw$) and $|f_w| \leqslant an$ (respectively, $|d_w| \leqslant an$).*

Proof. We will show the claim for D, the proof for E being symmetric. Choose $m \in \Gamma^*$ such that $S \stackrel{*}{\Rightarrow} m$. If $d_1, d_2 \in D$, then there exist $w_1, w_2 \in \Gamma^*$ such that $S \stackrel{*}{\Rightarrow} d_1Sw_1$ and $S \stackrel{*}{\Rightarrow} d_2Sw_2$. By suitably applying $n+1$ times the derivations $S \stackrel{*}{\Rightarrow} d_1Sw_1$ $S \stackrel{*}{\Rightarrow} d_1Sw_1$ and $S \stackrel{*}{\Rightarrow} m$, we can obtain a derivation of a word of the kind wmf, where $f \in \langle w_1, w_2\rangle$ and $|f| \leqslant n \max(|e_1|, |e_2|)$. Then we have the claim, by setting $f_w = f$. $\square$

Corollary 2.3. *Let d_1, d_2 be as above. If $\langle d_1, d_2\rangle$ is free, then $L(G)$ has exponential growth.*

Proof. If $\langle d_1, d_2\rangle$ is free, there are 2^n distinct words of the kind wmf_w, with $|w|_{d_1,d_2} \leqslant n$, as the mapping $w \to wmf_w$ is one-to-one. Then the language $L(G)$ has exponential growth, since a word of the kind wmf_w, has length $2n\max(|d_1|, |d_2|) + |m| + 2\max(|e_1|, |e_2|) \leqslant 2an$, where a is a constant.

We will need the fact that, if $a, b \in \Gamma^*$ and the submonoid generated by a and b is not free, then there exists $v \in \Gamma^*$ such that $a = v^n$, $b = v^m$, which implies that one of the two must be a prefix of the other (see [7]). $\square$

Lemma 2.4. *Let G be a grammar which does not have exponential growth. Then, for every, $k \in [1, n]$, $\gamma_D(n) \leqslant n$ and $\gamma_E(n) \leqslant n$.*

Proof. Again, we will show the claim for D, since the proof for E is symmetric. Order $D=\{d_0,\dots d_i\dots\}$ in such a way that $i<j \Rightarrow |d_i|\leqslant|d_j|$. Let $d_i,d_j\in D$, with $i<j$. Now, $\langle d_i,d_j\rangle$ cannot be free, since otherwise, by Corollary 2.3, $L(G)$ would have exponential growth. Then, by the above, d_i is a prefix of d_j, and, in particular, $i=j\Rightarrow d_i=d_j$. Then for every k there is at most one element of D of length k and then D has linear growth. □

3. Grammars of polynomial growth

Lemma 3.1. *Let Γ be an alphabet let $X,Y\subseteq\Gamma^*$ and $a,b\in\Gamma^*$. We have*
1. $\gamma_{XY}(n)\leqslant\gamma_X(n)\gamma_Y(n)$.
2. $\gamma_{X|Y}(n)\leqslant\gamma_X(n)+\gamma_Y(n)$.
3. $\gamma_{aXb}(n+|a|+|b|)=\gamma_X(n)$.

Proof. 1 and 2 are obvious. Since the mapping $x\to axb$ is a bijection, we have 3. □

Corollary 3.2. *Let $r_1,\dots,r_{m+1}\in\Gamma^*$ and let $L_1,\dots,L_m\subseteq\Gamma^*$ be of polynomial growth. Then the language $r_1Lr_2\cdots r_mLr_{m+1}$ has polynomial growth.*

Definition 3.3. If $n=1$, set $\Sigma'=\emptyset$, and, if $n>1$, set $\Sigma'=\{S_2,\dots,S_n\}$. For $k\in[1,n]$, denote by G_k be the grammar whose set of nonterminal symbols is Σ', whose axiom is S_k, and whose rules are the rules of G of the kind $A\to\alpha$, $A\in\Sigma'\cup\{S_k\}$ and $\alpha\in(\Sigma'\cup\Gamma)^*$.

Definition 3.4. Denote by $L'(G)$ the subset of $L(G)$ consisting of the words $w\in\Gamma^*$ which admit a derivation $S\Rightarrow w_1\cdots\Rightarrow w_m=w$ such that $w_i\in(\Sigma'\cup\Gamma)^*$, $\forall w_i\in[1,m]$.

Remark 3.5. If $n=1$, $L'(G)$ is the finite language of the words of $L(G)$ that can be obtained by a one-step derivation.

Lemma 3.6. *Let $n>1$. If, for every $k\in[2,n]$, $L(G_k)$ has polynomial growth, then $L'(G)$ has polynomial growth.*

Proof. Let $R(S)$ be the set of the right sides of the rules of G of the form $\{S\to\alpha\}$, with $\alpha\in(\Sigma'\cup\Gamma)^*$. Set $|R(S)|=m$. For every $i\in[1,m]$ we can write $\alpha_i=r_{1,i}S_{k_{1,i}}\cdots r_{x_i,i}S_{k_{x_i,i}}$ $r_{x_i+1,i}$, where $r_{1,i},\dots,r_{x_{i}+1,i}\in\Gamma^*$ and $S_{k_{1,i}},\dots,S_{k_{x_i,i}}\in\Sigma'_k$. Now, if a word w belongs to $L'(G)$, then it admits a derivation whose first term belongs to $R(S)$, and in which the symbol S never occurs. Then we have

$$\begin{aligned}L'(G)\subseteq\ & r_{1,1}L(G_{k_{1,1}})r_{2,1}L(G_{k_{2,1}})\cdots r_{x_1,1}L(G_{k_{x_1,1}})r_{x_1+1,1}\,|\\ &\cdots|r_{1,m}L(G_{k_{1,m}})r_{2,m}L(G_{k_{2,m}})\cdots r_{x_m,m}L(G_{k_{x_m,m}})r_{x_m+1,m},\end{aligned}$$

where the $k_{i,j}$ belong to $[2,n]$, and then, by the hypothesis, $L(G_{k_{i,j}})$ has polynomial growth. Then, by Corollary 3.2, we have the claim. □

4. The growth of $L(G)$

Definition 4.1. Let $w \in L(G)$. Set

$$P(w) = \{x \in DSE \,|\, \exists w_i, \dots, w_{i+m} = w \in (\Sigma' \cup \Gamma)^* \,| \\ S_k \Rightarrow w_1 \cdots \Rightarrow x \Rightarrow w_i \cdots \Rightarrow w_m = w\}.$$

Lemma 4.2. *Let $k \in [1,n]$. We have $L(G) \subseteq L'(G) \cup DL'(G)E$.*

Proof. Let $w \in L(G)$, and let $S \Rightarrow w_1 \cdots \Rightarrow w_m = w$ be a derivation of w. If $P(w) = \emptyset$, then, for every $i \in [1,m]$ we have $w_i \in (\Sigma' \cup \Gamma)^*$, so that $w \in L'(G)$. Otherwise, let $x \in P(w)$. There exist $\alpha, \beta \in (\Sigma' \cup \Gamma)^*$ such that $x = \alpha S \beta$, and a derivation $S \Rightarrow \cdots \Rightarrow x \Rightarrow w_i \Rightarrow, \dots \Rightarrow w_{i+m} = w$ with $w_i, \dots, w_{i+m} \in (\Sigma' \cup \Gamma)^*$. By changing the order of the derivation, we can always suppose that $\alpha, \beta \in \Gamma^*$. Then $\alpha \in D$, $\beta \in E$, and w has of the form $\alpha p \beta$, where $p \in L'(G)$. $\square$

Theorem 4.3. *Let $L(G)$ be a context-free language. Then its growth is either polynomial or exponential.*

Proof. We will show the result by induction on n, the number of non terminal symbols of G.

If $n = 1$, then we have the claim, by Lemmas 4.2, 2.4 and Remark 3.5. Let $n > 1$ and assume the theorem true for $L(G_k)$, for every $k \in [2,n]$, (G_k is a grammar on $n-1$ symbols). If $L(G_k)$ has polynomial growth, for every $k \in [2,n]$, then, by Lemma 3.6, L' also has polynomial growth. If not, by the induction hypothesis, there exists $k_0 \in [2,n]$ such that $L(G_{k_0})$ has exponential growth. Now, since G is proper, we have $a_{k_0}, b_{k_0} \in \Gamma^*$ and a derivation $S \overset{*}{\Rightarrow} a_{k_0} S_{k_0} b_{k_0}$. Then $a_{k_0} L(G_{k_0}) b_{k_0} \subseteq L(G)$ and $L(G)$ also has exponential growth. $\square$

5. Remarks

Two question naturally arise: to characterize the context-free languages of polynomial growth and to find an algorithm to decide whether a context-free language has polynomial or exponential growth.

6. Note added

After this paper was accepted for publication, M. Bridson and R. Gilman sent us a recent preprint, where they remark that Theorem 4.3 can be proved as a consequence of a result on grammars of sub-exponential growth in [1], and give a new proof of this last result based on a paper of S. Ginsburg and E. Spanier [5].

Acknowledgements

The author wishes to thank J.-P. Allouche for the careful revising of the article and his many helpful suggestions.

References

[1] M. Bridson, R. Gilman, Formal language theory and the geometry of 3-manifolds, Comment. Math. Helv. 71 (1996) 525, 555.
[2] M. Bridson, R. Gilman, Context-free languages of sub-exponential growth, preprint.
[3] N. Chomsky, M.P. Schützenberger, The algebraic theory of context-free languages, in: P. Braffort, D. Hirschberg (Eds.), Computer Programming and Formal Systems, North-Holland, Amsterdam, 1963, pp. 118–161.
[4] P. Flajolet, Analytic models and ambiguity of context-free languages, Theoret. Comput. Sci. 49 (1987) 283–309.
[5] S. Ginsburg, E. Spanier, Bounded ALGOL-like languages, Trans. Amer. Math. Soc. 113 (1964) 333–368.
[6] R.I. Grigorchuk, A. Machì, An example of indexed language of intermediate growth, Theoret. Comput. Sci. 215 (1999) 325–327.
[7] LOTHAIRE, Combinatorics on words. Advanced Book program, World Science Division, Addison-Wesley, Reading, MA, 1983.