ELSEVIER

Discrete Applied Mathematics 73 (1997) 1–4

DISCRETE APPLIED MATHEMATICS

Note
A decision method for Parikh slenderness of context-free languages

Juha Honkala *

Department of Mathematics, University of Turku, SF-20500 Turku, Finland

Received 15 January 1996; revised 15 March 1996

Abstract

In a recent paper we introduced Parikh slender languages as a generalization of slender languages defined and studied by Andraşiu, Dassow, Păun and Salomaa. Results concerning Parikh slender languages can be applied in ambiguity proofs of context-free languages. In this paper an algorithm is presented for deciding whether or not a given context-free language is Parikh slender.

1. Introduction and definitions

Length considerations are often useful in language theory. For example, Flajolet [5] has shown that the inherent ambiguity of very many context-free languages can be deduced from the transcendentality of their generating functions. Other deep results based on length considerations are well known, e.g., in the theory of Lindenmayer systems (see [14]).

The notion of a Parikh slender language was introduced and studied in [7]. The idea is to count the words with the same Parikh vector. By definition, a language is Parikh slender if the number of words in the language with the same Parikh vector is bounded from above. More precisely, if $\Sigma = \{x_1, \ldots, x_m\}$ is an alphabet and $w \in \Sigma^*$ is a word, the Parikh vector $\psi(w)$ of w is defined by

$$\psi(w) = (\#_{x_1}(w), \ldots, \#_{x_m}(w)),$$

where $\#_x(w)$ is the number of the occurrences of the letter x in w. Now, a language $L \subseteq \Sigma^*$ is termed *Parikh slender* if there is a positive integer k such that for each $(i_1, \ldots, i_m) \in \mathbb{N}^m$ there are at most k words in L with the Parikh vector $(i_1, \ldots, i_m)$. For basic results concerning Parikh slender languages see [7]. We mention only that Parikh

* E-mail: jhonkala@sara.cc.utu.fi.

PII S0166-218X(96)00023-6

slender languages can be used to give a new simple proof of the result of Autebert et al. [2] concerning the inherent ambiguity of coprefix languages of infinite words.

The notion of a Parikh slender language is a generalization of the notion of a slender language due to Andraşiu et al. [1]. For slender languages see also [3, 8, 9, 10–13].

The purpose of this note is to prove the following theorem.

Theorem 1. *It is decidable whether or not a given context-free language is Parikh slender.*

The reader is referred to [6, 15] for the results concerning context-free languages used in the next section.

2. The decision method

We proceed to decide whether or not a given context-free language is Parikh slender.

Suppose $L \subseteq \Sigma^*$ is a context-free language. First, decide by the method of [6, Theorem 5.5.2] whether or not L is a bounded language. If L is not bounded, [7, Theorem 4.1] implies that L is not Parikh slender. We continue with the assumption that L is bounded. Then it is possible to find effectively words $w_1, \ldots, w_n \in \Sigma^*$ such that

$$L \subseteq w_1^* w_2^* \ldots w_n^*.$$

Next, consider a new alphabet $\Delta = \{a_1, \ldots, a_n\}$ and define the morphism $h : \Delta^* \to \Sigma^*$ by $h(a_i) = w_i$, $1 \leqslant i \leqslant n$. Denote $R = a_1^* a_2^* \ldots a_n^*$. Clearly, $h(R) = w_1^* w_2^* \ldots w_n^*$. By the Cross-Section Theorem due to Eilenberg [4], there exists a rational language $R_1 \subseteq R$ such that h maps R_1 bijectively onto $w_1^* w_2^* \ldots w_n^*$. Furthermore, the construction of R_1 is effective. It follows that h maps $h^{-1}(L) \cap R_1$ bijectively onto L. Denote by ψ the Parikh mapping $\psi : \Delta^* \to \mathbb{N}^n$. Because $h^{-1}(L) \cap R_1$ is a context-free language $\psi(h^{-1}(L) \cap R_1)$ is an effectively obtainable semilinear set (see [15]). Suppose

$$\psi(h^{-1}(L) \cap R_1) = B_1 \cup \cdots \cup B_s, \tag{1}$$

where $B_i \subseteq \mathbb{N}^n$ is a linear set for $1 \leqslant i \leqslant s$.

Next, denote by ψ_1 the Parikh mapping $\psi_1 : \Sigma^* \to \mathbb{N}^m$ where m is the cardinality of Σ. To determine whether or not L is Parikh slender, we have to decide whether for every $k \geqslant 1$ there exist distinct words $v_1, \ldots, v_k \in h^{-1}(L) \cap R_1$ such that

$$\psi_1(h(v_1)) = \cdots = \psi_1(h(v_k)). \tag{2}$$

Because no two elements of $h^{-1}(L) \cap R_1$ have the same Parikh vector, it suffices to decide whether for every $k \geqslant 1$ there exist words $v_1, \ldots, v_k \in h^{-1}(L) \cap R_1$ with pairwise distinct Parikh vectors such that (2) holds.

Consider now the linear space $\mathbb{Q}^n$ over $\mathbb{Q}$ and define the linear mapping $\phi : \mathbb{Q}^n \to \mathbb{Q}^m$ by $\phi(e_i) = \psi_1(h(a_i))$ for $1 \leqslant i \leqslant n$. Here $\{e_1, \ldots, e_n\}$ is the natural basis of $\mathbb{Q}^n$. By the previous paragraph, it suffices to decide whether or not for every $k \geqslant 1$, there exist distinct vectors $x_1,\dots,x_k \in \psi(h^{-1}(L)\cap R_1)$ such that $\phi(x_1)=\cdots=\phi(x_k)$. By (1) it suffices to do this for each B_i, $1\leqslant i\leqslant s$. Consider, e.g., the set B_1 and suppose that

$$B_1=\left\{ b_0+\sum_{j=1}^{t} n_j b_j \mid n_j\in\mathbb{N} \text{ for } 1\leqslant j\leqslant t \right\},$$

where $b_j\in\mathbb{N}^n$ for $0\leqslant j\leqslant t$. Define the subspaces S_1 and S_2 of $\mathbb{Q}^n$ and $\mathbb{Q}^m$, respectively, by

$$S_1=\{q_1b_1+\cdots+q_tb_t \mid q_j\in\mathbb{Q} \text{ for } 1\leqslant j\leqslant t\}$$

and

$$S_2=\{q_1\phi(b_1)+\cdots+q_t\phi(b_t) \mid q_j\in\mathbb{Q} \text{ for } 1\leqslant j\leqslant t\}.$$

By elementary results from linear algebra, the dimensions of S_1 and S_2 can be computed effectively. If $\dim(S_1)=\dim(S_2)$, ϕ maps S_1 bijectively onto S_2 and there does not exist distinct vectors $x_1,x_2\in B_1$ such that $\phi(x_1)=\phi(x_2)$. If $\dim(S_1)>\dim(S_2)$, there exists a nonzero $x\in S_1$ such that $\phi(x)=0$. Without restriction we assume that

$$x=(n_1-m_1)b_1+\cdots+(n_t-m_t)b_t,$$

where $n_j, m_j\in\mathbb{N}$ for $1\leqslant j\leqslant t$. Denote

$$\alpha(k,q)=b_0+q(n_1b_1+\cdots+n_tb_t)+(k-q)(m_1b_1+\cdots+m_tb_t)$$

for nonnegative integers $k\geqslant 1$, $0\leqslant q\leqslant k$. Clearly, $\alpha(k,q)\in B_1$ for $k\geqslant 1$, $0\leqslant q\leqslant k$ and

$$\begin{aligned}\phi(\alpha(k,q)) &= \phi(b_0+k(m_1b_1+\cdots+m_tb_t))+\phi(qx)\\ &= \phi(b_0+k(m_1b_1+\cdots+m_tb_t))\end{aligned}$$

for $0\leqslant q\leqslant k$. Therefore, for every $k\geqslant 1$, there exist distinct $x_1,\dots,x_{k+1}\in B_1$ such that $\phi(x_1)=\cdots=\phi(x_{k+1})$.

We have now seen that it can be decided whether or not for every $k\geqslant 1$ there exist distinct $x_1,\dots,x_k\in\psi(h^{-1}(L)\cap R_1)$ such that $\phi(x_1)=\cdots=\phi(x_k)$. Therefore, it can be decided whether or not L is Parikh slender.

References

[1] M. Andraşiu, J. Dassow, G. Păun and A. Salomaa, Language-theoretic problems arising from Richelieu cryptosystems, Theoret. Comput. Sci. 116 (1993) 339–357.

[2] J.-M. Autebert, P. Flajolet and J. Gabarro, Prefixes of infinite words and ambiguous context-free languages, Inform. Process. Lett. 25 (1987) 211–216.

[3] J. Dassow, G. Păun and A. Salomaa, On thinness and slenderness of L languages, EATCS Bulletin 49 (1993) 152–158.

[4] S. Eilenberg, Automata, Languages and Machines, Vol. A (Academic Press, New York, 1974).
[5] P. Flajolet, Analytic models and ambiguity of context-free languages, Theoret. Comput. Sci. 49 (1987) 283–309.
[6] S. Ginsburg, The Mathematical Theory of Context-Free Languages (McGraw-Hill, New York, 1966).
[7] J. Honkala, On Parikh slender languages and power series, JCSS 52 (1996) 185–190.
[8] L. Ilie, On a conjecture about slender context-free languages, Theoret. Comput. Sci. 132 (1994) 427–434.
[9] T. Nishida and A. Salomaa, Slender 0L languages, Theoret. Comput. Sci. 158 (1996) 161–176.
[10] G. Păun and A. Salomaa, Decision problems concerning the thinness of D0L languages, EATCS Bulletin 46 (1992) 171–181.
[11] G. Păun and A. Salomaa, Closure properties of slender languages, Theoret. Comput. Sci. 120 (1993) 293–301.
[12] G. Păun and A. Salomaa, Thin and slender languages, Discrete Appl. Math. 61 (1995) 257–270.
[13] D. Raz, Length considerations in context-free languages, Theoret. Comput. Sci., to appear.
[14] G. Rozenberg and A. Salomaa, The Mathematical Theory of L Systems (Academic Press, New York, 1980).
[15] A. Salomaa, Formal Languages (Academic Press, New York, 1973).