Theoretical Computer Science 412 (2011) 5400–5408

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

The growth function of S-recognizable sets

Émilie Charlier [a,*], Narad Rampersad [b]

[a] David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada

[b] Institute of Mathematics, University of Liège, Grande Traverse 12 (B 37), B-4000 Liège, Belgium

ARTICLE INFO

Article history:
Received 30 December 2010
Accepted 27 May 2011
Communicated by D. Perrin

Keywords:
S-recognizable set
Growth function
Finite automaton
Regular language
Morphic sequence

ABSTRACT

A set $X \subseteq \mathbb{N}$ is S-recognizable for an abstract numeration system S, if the set $\mathrm{rep}_S(X)$ of its representations is accepted by a finite automaton. We show that the growth function of an S-recognizable set is always either $\Theta((\log(n))^{c-df} n^f)$ where $c, d \in \mathbb{N}$ and $f \geq 1$, or $\Theta(n^r \theta^{\Theta(n^q)})$, where $r, q \in \mathbb{Q}$ with $q \leq 1$. If the number of words of length n in the numeration language is bounded by a polynomial, then the growth function of an S-recognizable set is $\Theta(n^r)$, where $r \in \mathbb{Q}$ with $r \geq 1$. Furthermore, for every $r \in \mathbb{Q}$ with $r \geq 1$, we can provide an abstract numeration system S built on a polynomial language and an S-recognizable set such that the growth function of X is $\Theta(n^r)$. For all positive integers k and ℓ, we can also provide an abstract numeration system S built on an exponential language and an S-recognizable set such that the growth function of X is $\Theta((\log(n))^k n^\ell)$.

1. Introduction

A set $X \subseteq \mathbb{N}$ is *b-recognizable* if the set of representations of the elements of X in base b is accepted by a finite automaton. The class of b-recognizable sets is a very well-studied class (the main results can be found in the book of Allouche and Shallit [1]). It is therefore somewhat surprising that, to the best of our knowledge, no characterization of the possible growths of b-recognizable sets is currently known. In this paper, we provide such a characterization for the much more general class of S-recognizable sets.

Let L be a language over an alphabet Σ. We assume that the letters $a_1, \ldots, a_\ell$ of Σ are ordered by $a_1 < \cdots < a_\ell$. This order on the alphabet Σ induces an order $<$ on the language L called the *genealogic order* or the *radix order*: the words are ordered length by length, and for a given length, we use the lexicographic order. This leads to the definition of an abstract numeration system.

Definition 1. An *abstract numeration system* is a triple $S = (L, \Sigma, <)$, where L is an infinite language over a totally ordered finite alphabet $(\Sigma, <)$. The language L is called the *numeration language*. The map $rep_S : \mathbb{N} \to L$ is a bijection mapping $n \in \mathbb{N}$ to the $(n+1)$th word of L ordered genealogically. The inverse map is denoted by $val_S : L \to \mathbb{N}$.

Most of the time, we assume that L is a regular language. Nevertheless, some results hold true for arbitrary numeration languages.

The integer base numeration systems are particular cases of abstract numeration systems since for all $b, n, m \in \mathbb{N}$ with $b \geq 2$, we have $n < m \Leftrightarrow rep_b(n) < rep_b(m)$, where $rep_b(x)$ designates the usual greedy representation of x in base b. In this case, the numeration language is $rep_b(\mathbb{N}) = \{1, 2, \ldots, b-1\}\{0, 1, 2, \ldots, b-1\}^* \cup \{\varepsilon\}$, which is regular.

* Corresponding author. Tel.: +1 519 888 4567x33884; fax: +1 519 885 1208.
E-mail addresses: echarlier@uwaterloo.ca (É. Charlier), nrampersad@ulg.ac.be (N. Rampersad).

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2011.05.057

Definition 2. Let S be an abstract numeration system. Let $X \subseteq \mathbb{N}$. The set X is *S-recognizable* if the language $rep_S(X) = \{rep_S(n) : n \in X\}$ is regular. Let $b \geq 2$ be an integer. The set X is *b-recognizable* if it is S-recognizable for the abstract numeration system S built on the language $rep_b(\mathbb{N})$ consisting of the base-b representations of the elements of X. The set X is *1-recognizable* if it is S-recognizable for the abstract numeration system S built on a^*.

Note that $\mathbb{N}$ is S-recognizable if and only if the numeration language L is regular.

Rigo [11] stated the following two fundamental questions regarding S-recognizable sets:

- For a given numeration system S, what are the S-recognizable subsets of $\mathbb{N}$?
- For a given subset X of $\mathbb{N}$, is it S-recognizable for some numeration system S?

Of course, both of these questions are quite difficult to address in full generality. In this paper, we consider these questions in terms of the growth functions of both the set X and the numeration language.

Definition 3. For a subset X of $\mathbb{N}$, we let $t_X(n)$ denote the $(n+1)$th term of X. The map $t_X : \mathbb{N} \to \mathbb{N}$ is called the *growth function* of X.

We address the following problem.

Problem 1. Let S be an abstract numeration system built on a regular language. What do the growth functions of S-recognizable sets look like? Of course, this question has to be answered in terms of the growth function of the numeration language.

Apart from the following result, known as *Eilenberg's gap theorem*, we are not aware of any results in the direction of answering the above problem.

Theorem 4 *([8]). Let $b \geq 2$ be an integer. A b-recognizable set X of nonnegative integers satisfies either* $\limsup_{n\to+\infty}(t_X(n+1) - t_X(n)) < +\infty$ *or* $\limsup_{n\to+\infty} \frac{t_X(n+1)}{t_X(n)} > 1$.

Thanks to this result, examples of sets that are not b-recognizable for any b have been exhibited. The set $\{n^2 : n \in \mathbb{N}\}$ of squares is such an example. However, the set of squares is S-recognizable for the abstract numeration system

$$S = (a^*b^* \cup a^*c^*, \{a, b, c\}, a < b < c).$$

More generally, Rigo [11] and Strogalov [15] showed that for any polynomial $P \in \mathbb{Q}[x]$ such that $P(\mathbb{N}) \subseteq \mathbb{N}$, there exists S such that P is S-recognizable. Observe that in the case of an integer base numeration system, the number of words of each length in the numeration language grows exponentially, whereas in the case of the numeration system S, this number grows polynomially. This leads to the natural question: Can a set of the form $P(\mathbb{N})$ ever be recognized in a numeration system where the numeration language is exponential? In Section 6, we show that the answer to this question is no for all polynomials P of degree 2 or more.

Let us fix some asymptotic notation.

Definition 5. Let f and g be functions with domain $\mathbb{N}$. We say that f is $\Theta(g)$, and we write $f = \Theta(g)$, if there exist positive constants c and d and a nonnegative integer N such that, for all integers $n \geq N$, we have $cg(n) \leq f(n) \leq dg(n)$. We say that f and g have *equivalent behaviors at infinity*, which is denoted by $f(n) \sim g(n)$ $(n \to +\infty)$ (or simply $f \sim g$ when the context is clear), if we have $\lim_{n\to+\infty} \frac{f(n)}{g(n)} = 1$. Finally, we write $f = o(g)$ if we have $\lim_{n\to+\infty} \frac{f(n)}{g(n)} = 0$.

Definition 6. For any language L over an alphabet Σ and any nonnegative integer n, we let

$$\mathbf{u}_L(n) = |L \cap \Sigma^n|$$

denote the number of words of length n in L and

$$\mathbf{v}_L(n) = \sum_{i=0}^{n} \mathbf{u}_L(i) = |L \cap \Sigma^{\leq n}|$$

denote the number of words of length less than or equal to n in L. The maps $\mathbf{u}_L : \mathbb{N} \to \mathbb{N}$ and $\mathbf{v}_L : \mathbb{N} \to \mathbb{N}$ are called the *counting (or growth) functions of L*.

When L is a regular language, the sequence $(\mathbf{u}_L(n))_{n\geq 0}$ satisfies a linear recurrence relation with integer coefficients (for instance, see [2]): there exist a positive integer k and $a_1, \ldots, a_k \in \mathbb{Z}$ such that we have

$$\forall n \in \mathbb{N},\ \mathbf{u}_L(n+k) = a_1\mathbf{u}_L(n+k-1) + \cdots + a_k\mathbf{u}_L(n).$$

Then, since we have $\mathbf{v}_L(n) - \mathbf{v}_L(n-1) = \mathbf{u}_L(n)$, the sequence $(\mathbf{v}_L(n))_{n\geq 0}$ satisfies the linear recurrence relation of length $k+1$ whose characteristic polynomial is $(x-1)(x^k - a_1x^{k-1} - \cdots - a_k)$.

Our main result can be stated as follows.