Journal of Discrete Algorithms 14 (2012) 65–73

Contents lists available at SciVerse ScienceDirect

Journal of Discrete Algorithms

www.elsevier.com/locate/jda

Number of holes in unavoidable sets of partial words II ☆

F. Blanchet-Sadri [a,*], Steven Ji [b], Elizabeth Reiland [c]

[a] *Department of Computer Science, University of North Carolina, P.O. Box 26170, Greensboro, NC 27402-6170, USA*
[b] *Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*
[c] *Department of Mathematics, Harvey Mudd College, 301 Platt Boulevard, Claremont, CA 91711, USA*

ARTICLE INFO

Article history:
Available online 7 December 2011

Keywords:
Automata and formal languages
Computational complexity
Combinatorics on words
Partial words
Unavoidable sets
NP-hard problems

ABSTRACT

We are concerned with the complexity of deciding the avoidability of sets of partial words over an arbitrary alphabet. Towards this, we investigate the minimum size of unavoidable sets of partial words with a fixed number of holes. Additionally, we analyze the complexity of variations on the decision problem when placing restrictions on the number of holes and length of the words.

1. Introduction

An *unavoidable set of* (*full*) *words* X over an alphabet A is one such that any two-sided infinite word over A has a factor in X. Partial words, a generalization of full words, may contain "hole symbols", denoted by "$\diamond$'s", which are not considered part of the alphabet A. The $\diamond$ symbol is *compatible* with, or matches, each letter of A. An *unavoidable set of partial words* X over A is then defined as a set such that any two-sided infinite full word over A has a factor compatible with some element of X. This concept of unavoidable sets of partial words was introduced in [3].

Efficient algorithms to decide if a finite set X of full words over an alphabet A is unavoidable are well known [10]. For example, this check can be done by finding whether or not there is a loop in the automaton that recognizes $A^* \setminus A^*XA^*$, which must be finite for a set of words to be unavoidable [1]. This algorithm can be adapted to decide if a finite set X of partial words is unavoidable by determining the avoidability of $\hat{X}$, the completion of partial words in X. However, the computation is also much less efficient as a word with h holes can be completed in as many as $|A|^h$ ways. Avoidability, or the problem of deciding the avoidability of a finite set of partial words over a k-letter alphabet, where $k \geqslant 2$, turns out to be NP-hard [5,2], which is in contrast with the well-known feasibility results for a set of full words [7,10]. This can be proved by using a reduction from the 3SAT problem, known to be NP-complete. Avoidability also turns out to be in PSPACE [2].

In this paper, we prove several new results related to the complexity of deciding the avoidability of sets of partial words. More specifically, we calculate the minimum cardinality of unavoidable sets of partial words of a given length m with a fixed number of holes over a k-letter alphabet. Previous work has been done in the context of full words. Mykkeltveit, in

☆ This material is based upon work supported by the National Science Foundation under Grant No. DMS-0754154. The Department of Defense is also gratefully acknowledged. Part of this paper was presented at IWOCA 2010 (Blanchet-Sadri et al., 2011 [4]). A research assignment from the University of North Carolina at Greensboro for the first author is gratefully acknowledged. Some of this assignment was spent at the LIAFA: Laboratoire d'Informatique Algorithmique: Fondements et Applications of Université Paris 7-Denis Diderot, Paris, France.

* Corresponding author.
E-mail address: blanchet@uncg.edu (F. Blanchet-Sadri).

1570-8667/$ – see front matter
doi:10.1016/j.jda.2011.12.002

particular, showed that the minimum number of elements in an unavoidable set of full words of length m over an alphabet of size k is equal to $c(m,k)$, the number of conjugacy classes of words of that length over the given alphabet [12]. We also analyze the complexity of variations on the avoidability problem building on previous work by Blakeley et al. [2]. In particular, we study the complexity of deciding *aperiodic* (non-ultimately periodic) unavoidable sets of partial words. This notion, which is a natural extension of unavoidable sets, was introduced by Higgins and Saker in the context of full words [9]. In addition, we provide a new hard counting problem on partial words adding to previous work by Manea and Tiseanu [11].

The contents of our paper is as follows: In Section 2, we present the basic definitions and terminology regarding the major problem on unavoidable sets we are concerned with, that is, the *complexity problem* or the complexity of the problem of deciding the avoidability of a finite set of partial words over a k-letter alphabet. In Section 3, we provide some bounds on the minimum cardinality of unavoidable sets containing partial words of length m with h holes over a k-letter alphabet. In Section 4, we analyze the complexity of variations on the avoidability problem with restrictions put on the number of holes and length of the words. Additionally, we generalize the concept of aperiodic avoidability to sets of partial words and prove that the problem of deciding if a finite set of partial words over a k-letter alphabet is avoided by a one-sided aperiodic word is NP-hard. In Section 5, we present a hard counting problem on partial words. Finally in Section 6, we conclude with some remarks.

2. Preliminaries

Let A be a non-empty finite set called an *alphabet* whose elements we call *letters*. A *finite full word* (or simply finite word) w over A is a finite sequence of letters from A. We denote the length of w by $|w|$ and the $(i+1)$st letter of w by $w(i)$ (by convention, we index positions of w from zero). By ε we denote the empty word and by A^* the set of all finite words over A.

A *two-sided infinite full word* (or simply infinite word) w over A can be viewed as a function $w : \mathbb{Z} \to A$. We say that w has period p for some positive integer p, and call it p-periodic, if $w(i) = w(i+p)$ for all $i \in \mathbb{Z}$. If w has a period, we call it *periodic*. If v is a non-empty finite word, then we denote the unique infinite word $w = \cdots vvvvv \cdots$ such that $v = w(0)\cdots w(|v|-1)$ by $v^{\mathbb{Z}}$. Similarly, a *one-sided infinite full word* w can be viewed as a function $w : \mathbb{N} \to A$. We call w *ultimately periodic* if there exist finite words u and v ($v \neq \varepsilon$) such that $w = uvvvv\cdots$. We call a finite word v a *factor* of a word w if there exists some integer index i such that $v = w(i)\cdots w(i+|v|-1)$.

A *partial word* w of length m over A is a function $w : \{0,\ldots,m-1\} \to A_\diamond$ where $A_\diamond = A \cup \{\diamond\}$ with $\diamond \notin A$. The $\diamond$ symbol is referred to as a "hole". For the indices $0 \leqslant i \leqslant m-1$ such that $w(i) \in A$, we say that i is in the domain of w, denoted by $D(w)$. Otherwise, i is in the set of holes of w, denoted by $H(w)$. The set denoted by $A_\diamond^*$ represents the set of all finite words over $A_\diamond$ (i.e. the set of all finite partial words over A, including the empty word, ε). If a partial word can be written as $u_1 \diamond u_2 \diamond \cdots \diamond u_{n-1} \diamond u_n$, then the set $\{u_1 a_1 u_2 a_2 \cdots u_{n-1} a_{n-1} u_n \mid a_i \in A\}$ is a *partial expansion* on u. Note that the u_i's are not necessarily full words. In this paper, it is assumed, without loss of generality, that the first and last positions of every partial word in a set be defined (i.e. that these positions not be holes).

We say a finite partial word v is a *factor* of a partial word w if there exist x and y such that $w = xvy$. Two partial words u and v of equal length are said to be *compatible*, denoted as $u \uparrow v$, if $u(i) = v(i)$ for all $i \in D(u) \cap D(v)$. A word w is said to *meet* a set of partial words X if some element of X is compatible with a factor of w. A two-sided infinite word w *avoids* X if no factor of w is compatible with any element of X. If no two-sided infinite word avoids X, we say that X is *unavoidable*. Otherwise, we call X *avoidable*. In [3], an algorithm is given for deciding avoidability on the basis of four reductions that maintain avoidability: factoring, prefix–suffix, hole truncation and expansion. This reduction method will be used in some of our proofs.

Two full words u and v are said to be *conjugate* if there exist x and y such that $u = xy$ and $v = yx$. Conjugacy is an equivalence relation, which we can use to form equivalence classes of words of a given length m over a fixed alphabet of size k. The number of conjugacy classes is denoted by $c(m,k)$.

In the next sections, we examine some complexity problems on partial words related to Avoidability and some variations of it.

3. Minimum size of unavoidable sets of constant length

In [12], Mykkeltveit proved that for the case of full words, the minimal cardinality of an unavoidable set of words of constant length m over a k-letter alphabet, $\alpha(m,k)$, is precisely $c(m,k)$, the number of conjugacy classes of words of length m over a k-letter alphabet. The inequality $\alpha(m,k) \geqslant c(m,k)$ holds since an unavoidable set needs to contain at least one word from each conjugacy class. For example, if $m = 2$ and $k = 2$, there are three conjugacy classes $\{aa\}$, $\{bb\}$ and $\{ab, ba\}$ of words of length two over the binary alphabet $\{a,b\}$, and so $\{aa, bb, ab\}$ is an unavoidable set.

In this section, we are interested in the problem of calculating the cardinality of minimal unavoidable sets of partial words of length m with h holes over a k-letter alphabet, which we denote by $\alpha(m,h,k)$. Results, in the case of $h = 0$, have been obtained (for instance, see [13]). Using the algorithm for testing avoidability described in [3], Table 1 was obtained that gives $\alpha(m,h,k)$ for $2 \leqslant m \leqslant 10$, $0 \leqslant h \leqslant 8$, and $k = 2$. Note that an empty entry in the table indicates an impossible case (i.e. too many holes) or an entry that has not yet been discovered due to extensive computation time.