Information and Computation 241 (2015) 264–276

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Normality and two-way automata

Olivier Carton [a], Pablo Ariel Heiber [b,*]

[a] Laboratoire d'Informatique Algorithmique: Fondements et Applications, CNRS UMR 7089, Université Paris Diderot – Paris 7, Case 7014, 75205 Paris Cedex 13, France
[b] Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón I, Ciudad Universitaria, (1428) Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 15 April 2014
Received in revised form 20 November 2014
Available online 7 February 2015

Keywords:
Two-way automata
Normal numbers
Compression

ABSTRACT

We prove that two-way transducers (both deterministic and non-deterministic) cannot compress normal numbers. To achieve this, we first show that it is possible to generalize compressibility from one-way transducers to two-way transducers. These results extend a known result: normal infinite words are exactly those that cannot be compressed by lossless finite-state transducers, and, more generally, by bounded-to-one non-deterministic finite-state transducers. We also argue that such a generalization cannot be extended to two-way transducers with unbounded memory, even in the simple form of a single counter.

1. Introduction

Émile Borel defined normal numbers in 1909 [5]. A real number in the unit interval $[0, 1)$ is normal to an integer base b if all finite blocks of base b digits are uniformly distributed in its expansion in base b. Since normality to base b is a property of the expansion of real numbers in base b, it is natural to think of it as a property of infinite words over finite alphabets.

Even though for more than a hundred years there has been research on normality in several fields [6], some basic problems remain open. For instance, it is not known whether irrational constants such as π or e are normal to some base. This motivates the search for new characterizations of the concept of normality.

One characterization that has been around for some time is based on finite automata. An infinite word is normal if and only if it cannot be compressed by lossless finite transducers (also known as finite-state compressors). These are deterministic finite automata augmented with an output tape with injective input-output behavior. The compression ratio of an infinite run of a one-way transducer is defined as the $\liminf$, over all its finite prefixes, of the ratio between the amount of output made and input read up to that point. A given infinite word is said to be compressed by a given transducer if the compression ratio it achieves is less than 1.

A direct proof of the incompressibility characterization of normal words can be found in a recent paper of Becher and Heiber [4]. However, the result was already known, although by indirect and more involved arguments. For instance, combining results of Schnorr and Stimm [13] and Dai, Lathrop, Lutz and Mayordomo [7] yields an earlier proof.

Together with Becher, we investigated how much power can be added to these deterministic transducers while retaining incompressibility of normal infinite words, in terms of combinations of non-determinism and different models of unbounded memory [3]. The automata studied in that work are always one-way; that is, they process the input from left to right. This method of processing plays a major role in the definition of compressibility, even in the earlier papers on the subject. In the

* Corresponding author.
E-mail addresses: Olivier.Carton@liafa.univ-paris-diderot.fr (O. Carton), pheiber@dc.uba.ar (P.A. Heiber).

http://dx.doi.org/10.1016/j.ic.2015.02.001
0890-5401/

rest of the paper, these transducers are called one-way transducers to emphasize the difference with two-way transducers that we consider.

In this work we study the power of compression of two-way transducers, in particular regarding the compressibility of normal infinite words. As opposed to one-way machines considered in the past, two-way transducers have an input reading head that may be moved backwards to re-read parts of the input several times.

The introduction of two-way transducers can be traced back to the very beginning of the study of transducers [1]. Since then, it has been shown that the equivalence of deterministic such transducers is decidable [9]. In fact, decidability was shown to hold for the equivalence problem in a broader class of transducers: those that are single-valued, that is, functional [10]. These transducers also have a nice logical characterization [8]. The interest in these transducers has recently been renewed by the introduction of an equivalent model called streaming string transducers [2].

The definition of compressibility used in the past for one-way machines does not generalize in a unique way to two-way machines. Hence, a major part of the work we present is to consider several ways in which the definition may be generalized and prove that they are equivalent. The extent of this equivalence is stronger for deterministic two-way transducers than for non-deterministic ones. This implies that results for non-deterministic transducers are not strict generalizations of the results for the deterministic case, which motivates us to present them separately. However, a significant part of the work done for the deterministic case is reused for the case of non-deterministic transducers.

The main result of our work is that normal infinite words are not compressible by deterministic nor by non-deterministic two-way transducers. In the last section, we show with an example that trying to generalize to two-way transducers with unbounded memory the traditional definition of compressibility yields undesirable results: every infinite word turns out to be compressible by a two-way transducer with a single counter.

2. Preliminaries

2.1. Notation

If A is any finite set, we denote its cardinality by $|A|$. An alphabet is a finite set of symbols, and a finite or infinite word over a given alphabet A is a finite or infinite, respectively, sequence of elements of A. We denote by A^ℓ the set of words of length ℓ from A, and by $A^* = \bigcup_{\ell \geq 0} A^\ell$ the set of all words over A. We use uppercase letters to denote sets or other complex objects, lowercase letters a, b, c to denote symbols, u, v, w to denote finite words and x, y, z to denote infinite words. Usually, p, q denote elements of a set of states Q defined in the context. Given finite words u and v and an infinite word x over the same alphabet, $|u|$ is the length of u, uv is the concatenation of u and v, ux is the concatenation of u and x. For positions $m \leq n$, $u[m..n]$ and $x[m..n]$ are the words occurring from position m to position n, inclusive, in u and x respectively, and $x[m..]$ is the suffix of x starting at position m. We write log for the logarithm in base 2. Tuples are denoted with angle brackets and commas: a k-tuple is denoted by $\langle x_1, x_2, \ldots, x_k \rangle$. A function f is t-to-one if the function that maps each y to its number $|\{x : f(x) = y\}|$ of pre-images is bounded by t. A function is bounded-to-one if it is t-to-one for some integer t. Given two functions f and g we use the usual "little o" notation and say that $f \in o(g)$ whenever $\lim_{n\to\infty} f(n)/g(n) = 0$.

2.2. Normality

There are several equivalent definitions of normality. Here we give the one that is most convenient for our presentation. We refer the reader to Bugeaud's book [6] for a comprehensive list of possible definitions of normality and the proofs of their equivalence.

Definition 1. An infinite word x over an alphabet A is *ℓ-simply normal* if all words of length ℓ occur with the same asymptotic frequency in the blocks of size ℓ of x, namely, if $x = v_1 v_2 v_3 \cdots$ with $|v_i| = \ell$,

$$\forall u \in A^\ell, \quad \lim_{n\to\infty} \frac{|\{i \leq n : u = v_i\}|}{n} = |A|^{-\ell}.$$

An infinite word x is *normal* if it is ℓ-simply normal for all positive ℓ.

If the infinite word x is the expansion in some base $b \geq 2$ of some real number α, grouping the symbols in consecutive blocks of size ℓ corresponds to considering the expansion in base b^ℓ of α.

We make use of the following lemma which bounds the gap between consecutive occurrences of a word in the blocks of a normal infinite word.

Lemma 2.1. *Let x be a normal infinite word and w a word over the same alphabet A. There is an increasing function $h(n) \in o(n)$ depending on x and w, such that for any n there is an occurrence of w in the word $x[n..n + h(n)]$.*