Introducing Privileged Words: Privileged Complexity of Sturmian Words

Jarkko Peltomäki

University of Turku
TUCS Turku Centre for Computer Science

19.9.2013

Introduction

Privileged words are a new class of words introduced recently in

Kellendonk, Lenz, Savinien:
A Characterization of Subshifts with Bounded Powers (2011).

There are only a few papers written about these words. In this talk I introduce the concept of privileged words, and describe a characterization of Sturmian words using so-called privileged complexity function.

Before giving the definition let us see how the concept arises from the theory of rich words.

A word w is *rich* if it has exactly $|w| + 1$ distinct palindromic factors including the empty word (this is the maximum value that can be attained). An infinite word is rich if its every factor is rich.

Motivation

A word w is *rich* if it has exactly $|w| + 1$ distinct palindromic factors including the empty word (this is the maximum value that can be attained). An infinite word is rich if its every factor is rich.

Example

The word 0010110 is rich whereas 00101100 is not.

Motivation

A word w is *rich* if it has exactly $|w| + 1$ distinct palindromic factors including the empty word (this is the maximum value that can be attained). An infinite word is rich if its every factor is rich.

Example

The word 0010110 is rich whereas 00101100 is not.

Definition

Word w is *a complete first return* to the word v if w begins and ends with v and has exactly two occurrences of v.

Motivation

Rich words have the following characterization:

Theorem (Glen et al., 2009)

A finite or infinite word w is rich if and only if every complete first return to a palindrome in w is itself a palindrome.

Rich words have the following characterization:

Theorem (Glen et al., 2009)

A finite or infinite word w is rich if and only if every complete first return to a palindrome in w is itself a palindrome.

This can be interpreted as "words of class X are complete first returns to shorter words of class X" where X is the class of palindromes, and the shortest palindromes are the empty word and the letters.

Borrowing the previous idea we define the set of *privileged words* over the alphabet A, denoted $\mathcal{P}ri(A)$, recursively as follows:

Definition

Word w, $|w| \geq 2$, is privileged if it is a complete first return to a shorter privileged word. The shortest privileged words are the empty word and the letters of the alphabet A.

Borrowing the previous idea we define the set of *privileged words* over the alphabet A, denoted $\mathcal{P}ri(A)$, recursively as follows:

Definition

Word w, $|w| \geq 2$, is privileged if it is a complete first return to a shorter privileged word. The shortest privileged words are the empty word and the letters of the alphabet A.

Example

$\mathcal{P}ri(\{0,1\}) = \{\varepsilon, 0, 1, 00, 11, 010, 101, \ldots, 00101100, \ldots\}$

Definition

Borrowing the previous idea we define the set of *privileged words* over the alphabet A, denoted $\mathcal{P}ri(A)$, recursively as follows:

Definition

Word w, $|w| \geq 2$, is privileged if it is a complete first return to a shorter privileged word. The shortest privileged words are the empty word and the letters of the alphabet A.

Example

$\mathcal{P}ri(\{0,1\}) = \{\varepsilon, 0, 1, 00, 11, 010, 101, \ldots, 00101100, \ldots\}$

Not every palindrome is privileged: 0120210.

Lemma

Let w be a privileged word, and u its any privileged prefix (suffix). Then u is a suffix (prefix) of w.

Basic Properties

Lemma

Let w be a privileged word, and u its any privileged prefix (suffix). Then u is a suffix (prefix) of w.

Lemma

Any word w has exactly $|w| + 1$ distinct privileged factors. That is, every word is "rich" in privileged words.

At first it seems that privileged words have nothing to do with palindromes. However the following can be proved:

At first it seems that privileged words have nothing to do with palindromes. However the following can be proved:

Proposition

Let w be a finite or infinite word. Then w is rich if and only if $\mathcal{P}ri(w) = \mathcal{P}al(w)$.

Relation to Rich Words

At first it seems that privileged words have nothing to do with palindromes. However the following can be proved:

Proposition

Let w be a finite or infinite word. Then w is rich if and only if $\mathcal{Pri}(w) = \mathcal{Pal}(w)$.

This result is a straightforward consequence of the fact that in rich words complete first returns to palindromes are palindromes.

Some Complexity Functions

In combinatorics on words many so-called complexity functions have been studied.

In combinatorics on words many so-called complexity functions have been studied.

The Factor Complexity Function

The factor complexity function $\mathcal{C}_w(n)$ counts the number of distinct factors of length n in the word w. That is $\mathcal{C}_w(n) = |\mathcal{F}_w(n)|$.

In combinatorics on words many so-called complexity functions have been studied.

The Factor Complexity Function

The factor complexity function $\mathcal{C}_w(n)$ counts the number of distinct factors of length n in the word w. That is $\mathcal{C}_w(n) = |\mathcal{F}_w(n)|$.

The Palindromic Complexity Function

The palindromic complexity function $\mathcal{P}_w(n)$ counts the number of distinct palindromes of length n in the word w. That is $\mathcal{P}_w(n) = |\mathcal{P}al_w(n)|$.

Some Complexity Functions

The Privileged Complexity Function

Similarly by setting $\mathcal{A}_w(n) = |\mathcal{P}ri_w(n)|$ we obtain the privileged complexity function of w which counts the number of privileged factors of length n in w.

The Privileged Complexity Function

Similarly by setting $\mathcal{A}_w(n) = |\mathcal{P}ri_w(n)|$ we obtain the privileged complexity function of w which counts the number of privileged factors of length n in w.

In rich words $\mathcal{P}_w(n) = \mathcal{A}_w(n)$ for every $n \geq 0$ as in rich words palindromes and privileged factors are the same thing.

The Privileged Complexity Function

Similarly by setting $\mathcal{A}_w(n) = |\mathcal{Pri}_w(n)|$ we obtain the privileged complexity function of w which counts the number of privileged factors of length n in w.

In rich words $\mathcal{P}_w(n) = \mathcal{A}_w(n)$ for every $n \geq 0$ as in rich words palindromes and privileged factors are the same thing.

Actually rich words are characterized by this property.

Definition

An infinite word w is Sturmian if $\mathcal{C}_w(n) = n + 1$ for all $n \geq 0$.

Definition

An infinite word w is Sturmian if $\mathcal{C}_w(n) = n + 1$ for all $n \geq 0$.

Fibonacci Word

The fixed point $0100101001001010\ldots$ of the morphism $0 \to 01, 1 \to 0$ is Sturmian.

Definition

An infinite word w is Sturmian if $\mathcal{C}_w(n) = n + 1$ for all $n \geq 0$.

Fibonacci Word

The fixed point $0100101001001010\ldots$ of the morphism $0 \to 01, 1 \to 0$ is Sturmian.

Sturmian words have a large number of different characterizations, many of them are in terms of different complexity functions.

Main Result

The palindromic complexity function of Sturmian words completely describes Sturmian words.

Theorem (Droubay and Pirillo, 1999)

An infinite word w is Sturmian if and only if

$$\mathcal{P}_w(n) = \begin{cases} 1, & \text{if } n \text{ is even,} \\ 2, & \text{if } n \text{ is odd.} \end{cases}$$

Main Result

The palindromic complexity function of Sturmian words completely describes Sturmian words.

Theorem (Droubay and Pirillo, 1999)

An infinite word w is Sturmian if and only if

$$\mathcal{P}_w(n) = \begin{cases} 1, & \text{if } n \text{ is even,} \\ 2, & \text{if } n \text{ is odd.} \end{cases}$$

The main result is that a similar characterization can be given in terms of the privileged complexity function.

Theorem

An infinite word w is Sturmian if and only if

$$\mathcal{A}_w(n) = \begin{cases} 1, & \text{if } n \text{ is even,} \\ 2, & \text{if } n \text{ is odd.} \end{cases}$$

Episturmian words and codings of interval exchange transformations (IET words) are two generalizations of Sturmian words to larger alphabets. Episturmian words and IET words have the same palindromic complexity and privileged complexity, but they are not the same class of words. Hence the main result doesn't generalize to larger alphabets: privileged complexity can't distinguish episturmian words and IET words in general.

In rich words palindromes are the same thing as privileged factors. Thus privileged words are not really useful when studying rich words.

In rich words palindromes are the same thing as privileged factors. Thus privileged words are not really useful when studying rich words.

However in non-rich words there are always many privileged factors, but not that many palindromic factors. In a non-rich word privileged words may behave radically differently than palindromes.

The Thue-Morse Word

Take as an example the Thue-Morse word

$$t = 0110100110010110 1001\ldots,$$

a fixed point of the morphism $0 \to 01, 1 \to 10$.

The Thue-Morse Word

Take as an example the Thue-Morse word

$$t = 0110100110010110 1001\ldots,$$

a fixed point of the morphism $0 \to 01, 1 \to 10$.

Theorem (Allouche et al., 2003 & Blondin-Massé et al., 2008)

The palindromic complexity function $\mathcal{P}_t(n)$ satisfies

$$\mathcal{P}_t(0) = 1, \mathcal{P}_t(1) = \mathcal{P}_t(2) = \mathcal{P}_t(3) = \mathcal{P}_t(4) = 2,$$
$$\mathcal{P}_t(4n) = \mathcal{P}_t(4n-2) = \mathcal{P}_t(n) + \mathcal{P}_t(n+1), n \geq 2,$$
$$\mathcal{P}_t(2n+1) = 0, n \geq 2.$$

We denote by $\mathcal{A}_u(n)$ the number of privileged factors of length n of t which begin with u.

We denote by $\mathcal{A}_u(n)$ the number of privileged factors of length n of t which begin with u.

Theorem

The privileged complexity function $\mathcal{A}_t(n)$ satisfies

$$\begin{aligned}
\mathcal{A}_t(0) &= 1, \mathcal{A}_t(1) = \mathcal{A}_t(2) = \mathcal{A}_t(3) = \mathcal{A}_t(4) = 2, \\
\frac{1}{2}\mathcal{A}_t(4n) &= 3\mathcal{A}_{00}(n) + \mathcal{A}_{010}(n) \\
&\quad + \mathcal{A}_{010}(n+1) + \mathcal{A}_{0110}(n+1), n \geq 2, \\
\frac{1}{2}\mathcal{A}_t(4n-2) &= \mathcal{A}_{00}(4(n-1)) + \mathcal{A}_{010}(4n) + \mathcal{A}_{0110}(4n), n \geq 2, \\
\mathcal{A}_t(2n+1) &= 0, n \geq 2.
\end{aligned}$$

Using the previous formulas it can be proven that the function $\mathcal{A}_t(n)$ is unbounded.

Using the previous formulas it can be proven that the function $\mathcal{A}_t(n)$ is unbounded.

This differs radically from palindromes: the palindromic complexity function of a fixed point of a primitive morphism is bounded.

Thank You

Thank you for your attention!

Jarkko Peltomäki
Introducing Privileged Words: Privileged Complexity of Sturmian Words
Theoretical Computer Science Vol. 500 (2013)

Jarkko Peltomäki
Privileged Factors in the Thue-Morse Word – A Comparison of Privileged Words and Palindromes
arXiv:1306.6768