Applied Mathematics and Computation 329 (2018) 395–401

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

Factor complexity and permutation complexity of the generalized Morse sequence

Min Niu*, Qiuxia Liu

Department of Applied Mathematics, University of Science and Technology Beijing, Beijing 100083, PR China

ARTICLE INFO

MSC:
68R15
11B85

Keywords:
Generalized Morse sequence
Factor complexity
Permutation complexity

ABSTRACT

Some properties of the generalized Morse sequence and the permutation associated with it are considered. Factor complexity and permutation complexity formulas of it are established.

1. Introduction

Infinite permutations and permutation complexity of infinite words attract more and more researchers' interest in recent years. For the cases of the period doubling sequence, the Sturmian sequences and the Thue–Morse sequence and so on, the reader is directed to refer Refs. [1,3–9] for more information. While the factor complexity of the above sequences have been studied much earlier.

In this paper, we are interested in the study of the generalized Morse sequence. Before stating our main interest, let us first recall some basic notations and definitions [2,3,6,9].

Let $\mathscr{A}$ be a finite alphabet set, a word is a finite string of elements in $\mathscr{A}$ and the set of all finite words over $\mathscr{A}$ is denoted by $\mathscr{A}^*$. The empty word is denoted by ε. Let $\mathscr{A}^N$ be the set of sequences (or infinite words) over $\mathscr{A}$.

The concatenation of two words $u = u_0 \cdots u_n$ and $v = v_0 \cdots v_m$ is the word $uv = u_0, \ldots, u_n v_0, \ldots, v_m$. $u[i, j] = u_i, \ldots, u_j$ or $u[i] = u_i, \ldots, u_n$ is said to occur at position i in the word u. And $u[i, j]$ also denotes a factor (or subword) of u, denoted by $u[i, j] \prec u$. Word $u[0, i]$ is called the prefix of u and $u[i, n]$ the suffix of u, written $u[0, i] \lhd u$ and $u[i, n] \rhd u$, respectively. A subword u of a given sequence $w \in \mathscr{A}^N$ is called the special word if there exist at least two distinct letters $a, b \in \mathscr{A}$ such that ua, ub are still subwords of w.

The number of letters appearing in a word u is called the length of u, denoted by $|u|$. The number of distinct words of length s occurring in u is called the factor complexity function of u, and denoted by $p_u(s)$. In a natural way, these definitions can be extended to infinite words.

We denote a sequence $(u_n)_{n\in N} = u_0 u_1, \ldots, u_n, \ldots$. Let T denote the following map defined on $\mathscr{A}^N$, called the one-sided shift:

$$T((u_n)_{n\in N}) = (u_{n+1})_{n\in N}.$$

In addition, by induction, we set $T^i((u_n)_{n\in N}) = (u_{n+i})_{n\in N}$ for $i = 2, 3, \cdots$, and by convention, $T^0((u_n)_{n\in N}) = (u_n)_{n\in N}$, respectively.

* Corresponding author.
E-mail address: niuminfly@sohu.com (M. Niu).

https://doi.org/10.1016/j.amc.2018.02.019
0096-3003/

A morphism $\sigma : \mathscr{A}^* \to \mathscr{A}^*$ is called a substitution of $\mathscr{A}^*$. And σ satisfies $\sigma(\varepsilon) = \varepsilon$ and $\sigma(uv) = \sigma(u)\sigma(v),\ u,\ v \in \mathscr{A}^*$. In this paper, we will mainly consider $\mathscr{A} = \{0, 1\}$.

The symbol a^n, $n \geq 1$, $a \in \mathscr{A}$ is an abbreviation for the word

$$\overbrace{aa\cdots a}^{n}.$$

And a^0 is equal to the empty word. We use the notation $\overline{a} = 1 - a$.

Let the integer $m \geq 2$ be fixed. The generalized Morse sequence w is generated by the morphism σ defined as: $0 \to 01^m$, $1 \to 10^m$. We have $w = \sigma^\infty(0) := \lim_{n\to\infty} \sigma^n(0)$, that is,

$$w = 01^m(10^m)^m10^m \ldots.$$

It is known that w is a fixed point of σ, that is $w = \sigma(w)$. An n-word is any one of the words $\sigma^n(a)$, $a \in \{0, 1\}$. Let A_n be an n-word beginning with 0 and B_n an n-word beginning with 1, that is $A_n = \sigma^n(0)$ and $B_n = \sigma^n(1)$.

A permutation π is a triple $\pi = \langle X, <, \prec_\pi \rangle$, where X is a countable set, $<$ and $\prec_\pi$ are total orders on X. We recall that for any aperiodic infinite word $w = w_0w_1, \ldots, w_n, \ldots$, on the binary alphabet $\mathscr{A} = \{0, 1\}$, no two shifts on w are equal lexicographically. For any aperiodic infinite word w and any $i, j \in X = \mathbb{N} \cup \{0\}$, we say $i \prec_{\pi_w} j$ if $w[i] < w[j]$, else we say $j \prec_{\pi_w} i$. While $w[i] < w[j]$ denotes the lexicographical order on sequences induced by $0 < 1$. Now we call π_w is an infinite permutation associated with w.

For any $n \geq 1$ consider the factor $w[a, a + n - 1]$ of w with length n, let $\pi_w[a, a + n - 1]$ be the finite permutation of $\{1, 2, \ldots, n\}$ such that for each $0 \leq i, j \leq n - 1$, $\pi_w[a, a + n - 1](i) < \pi_w[a, a + n - 1](j)$ if and only if $w[a + i] < w[a + j]$. We say $p = p_0p_1, \ldots, p_{n-1}$ is a subpermutation of π_w of length n if $p = \pi_w[a, a + n - 1]$. Finally, the permutation complexity function of w is defined as the number of different subpermutation of π_w of length n, which is denoted by $\tau_w(n)$.

Let us now give an example to explain the definitions of permutation and permutation complexity for the reader's convenience.

Example 1.1. We consider the well-known Thue–Morse sequence

$$t = 0110100110010110\cdots\cdots$$

- Let $u = 100 = t[4, 6] = t[8, 10]$. Then we have $\pi_t[4, 6] = \pi_t[8, 10] = (312)$ since $t[4] > t[6] > t[5]$ and $t[8] > t[10] > t[9]$.
- Let $v = 010 = t[3, 5] = t[10, 12]$. Then we have $\pi_t[3, 5] = (231)$ and $\pi_t[10, 12] = (132)$. We see that the same word 010 has different permutations.

In the same way, we have $\pi_t[1, 3] = (321)$, $\pi_t[5, 7] = (123)$ and $\pi_t[11, 13] = (213)$. Now we can see that for the word with length three of the Thue–Morse sequence, there are at least six different permutations. The trivial upper bound for $\tau_t(3)$ being the number of permutations of length 3, which is 3!. Then we can get $\tau_t(3) = 6$.

The permutation complexity of the Thue–Morse sequence is studied by Widmer [9] and the following result is obtained: for any $n \geq 2$, where $n = 2^a + b$ with $0 < b \leq 2^a$, the permutation complexity $\tau_t(n) = 2(2^{a+1} + b - 2)$. We remark that the proof of the permutation complexity of the Thue–Morse sequence in [9] does not use the factor complexity function. However, our calculation of the permutation complexity of the generalized Morse sequence will rely on the factor complexity. Here we point out that the special words play an important role in computation of the factor complexity and the permutation complexity of the generalized Morse sequence.

This paper is organized as follows: In the second section, we study some properties of the generalized Morse sequence. According to these properties, factor complexity formula is given. The third section is devoted to investigating properties of the permutation associated with the generalized Morse sequence. Finally, we work out the permutation complexity of the generalized Morse sequence. Low order subpermutations are listed in Appendix to be used as a base case for induction arguments.

2. Combinatorial properties of the generalized Morse sequence

In this section, we will study the factor structure and the factor complexity of the generalized Morse sequence.

Lemma 2.1. *For the generalized Morse sequence, we have* $w_{(m+1)n} = w_n,\ w_{(m+1)n+i} = \overline{w_n},\ i \in [1, m]$.

Proof. Since w is a fixed point of σ, that is $w = \sigma(w) = \sigma(w_0)\sigma(w_1), \ldots, \sigma(w_n), \ldots,$ we have $w_{(m+1)n}w_{(m+1)n+1}, \ldots, w_{(m+1)n+m} = \sigma(w_n)$. Thus we have $w_{(m+1)n} = w_n,\ w_{(m+1)n+i} = \overline{w_n},\ i \in [1, m]$. □

Lemma 2.2. *The generalized Morse sequence is aperiodic.*

Proof. From Lemma 2.1, we know that $w_{(m+1)n+i} = \overline{w_n},\ i \in [1, m]$. Therefore $m \cdot n + i,\ i \in [1, m]$ are not period. For any natural number p, there always exist some n and i such that $p = m \cdot n + i$. □

In the following we should use the word bar. A bar is used to divide a factor into 1-words.

Lemma 2.3. *In the generalized Morse sequence, every factor with length at least* $m + 2$ *has a unique decomposition into 1-words.*