Pattern Avoidance and HDOL Words

Pascal Ochem

▶ To cite this version:

Pascal Ochem. Pattern Avoidance and HDOL Words. WORDS 2011 - 8th International Conference on Combinatorics on Words, Sep 2011, Prague, Czech Republic. , 2011. lirmm-00626116

HAL Id: lirmm-00626116

https://hal-lirmm.ccsd.cnrs.fr/lirmm-00626116v1

Submitted on 8 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pattern Avoidance and HDOL Words

Pascal Ochem

LIRMM, CNRS, Montpellier

WORDS 2011

Patterns

Pattern: finite word over $\{A, B, C, \cdots\}$.
Word: over $\Sigma_k = \{0, 1, \cdots, k-1\}$.
Occurrence: obtained from the pattern by replacing each letter by a non-empty word.

Example

- 10111011 is an occurrence of ABCACB with $A = 10$, $B = 1$, $C = 1$.
- 00101001 contains 3 occurrences of ABBA: 001010, 010100 and 1001.

w avoids P if w contains no occurrence of P as a factor.
Avoidability index: $\lambda(P)$ is the smallest k such that there exists an infinite word over Σ_k avoiding P.

Avoidability index (1)

- $\lambda(AA) = 3$: infinite ternary square-free words.
- $\lambda(AAA) = 2$: infinite binary cube-free words.

Remark: $\lambda(P) \geq 2$

Divisibility:
ABCABCAAC contains an occurrence *AABB*
$\Longrightarrow \lambda(ABCABCAAC) \leq \lambda(AABB)$
$\Longrightarrow$ partial order among patterns

Avoidability index (2)

- $\cdots \leq \lambda(ABCABC) \leq \lambda(ABAB) \leq \lambda(AA) = 3$.
- $\cdots \leq \lambda(AABBCC) \leq \lambda(AABB) \leq \lambda(AA) = 3$.

We have:

- $\lambda(ABCABC) = 2$, $\lambda(ABAB) = 3$.
- $\lambda(AABBCC) = 2$, $\lambda(AABB) = 3$.

Bounds on the avoidability index

Lower bounds: To prove that $\lambda(P) \geq k$, check that all words over Σ_{k-1} are finite by lexicographic enumeration. This is the easy part.

Upper bounds: To prove that $\lambda(P) \leq k$, construct an infinite word over Σ_k and prove that it avoids P.

Square-free words and square-free morphisms

$\lambda(AA) = 3$: infinite ternary square-free word.

square-free morphisms:

- $0 \mapsto 01201$, $1 \mapsto 020121$, $2 \mapsto 0212021$.
- $0 \mapsto 0120212012102120210$,
  $1 \mapsto 1201020120210201021$,
  $2 \mapsto 2012101201021012102$.

More than 1.3^n ternary square-free words of size n.

Morphic words

Pure morphic word (DOL):
fixed point $m^\infty(a)$ of a morphism m such that $m(a) = aw$.
Morphic word (HDOL):
image $h(m^\infty(a))$ of a pure morphic word by a morphism h.

- $\lambda(AA) = 3$, avoided by pure morphic words.
- s: $0 \mapsto 0101$, $1 \mapsto 0011$, $2 \mapsto 1000$.
  The image by s of a square-free word avoids $ABCABC$.
- $\lambda(ABCABC) = 2$, avoided by a morphic word, but no pure morphic word.
- Exponentially many binary words avoid $ABCABC$.

Julien Cassaigne's conjecture:
$\lambda(P) = k$ if and only if there exists a morphic word over Σ_k avoiding P.

A square-free morphic word

m_1: $0 \mapsto 012$, $1 \mapsto 02$, $2 \mapsto 1$.

$m_1^\infty(0)$ is the only infinite ternary word avoiding

- squares, 010, and 212.
- squares and $0u1u0$ for $u \in \Sigma_3^*$.

m_1 is not square-free:
$m_1(0u1u0) = 012m_1(u)02m_1(u)012$ contains $(2m_1(u)0)^2$.

Characterization

- A set S of forbidden patterns and factors.
- A morphic word w over Σ_k.

S characterizes w if w avoids S and for every finite factor f of w, $S \cup f$ is unavoidable over Σ_k.

Example

- $\{AA, 010, 212\}$ characterizes $m_1^\infty(0)$.
- $\{ABABA, 000, 111\}$ characterizes the Thue-Morse word.

This notion is defined to handle extendability issues.
Makes no difference between a pattern and the associated formula.

Other characterizations by Thue (1)

m_2: $0 \mapsto 130402$, $1 \mapsto 132$, $2 \mapsto 1304$, $3 \mapsto 1402$, $4 \mapsto 1404$.

h_1: $0 \mapsto 012102120210120212$,
$1 \mapsto 01210212$,
$2 \mapsto 01210212021$,
$3 \mapsto 01210120212$,
$4 \mapsto 0121012021$.

h_2: $0 \mapsto 0210120102012$,
$1 \mapsto 021012$,
$2 \mapsto 02101201$,
$3 \mapsto 02102012$,
$4 \mapsto 0210201$.

- $\{AA, 010, 020\}$ characterizes $h_1(m^\infty(1))$
- $\{AA, 121, 212\}$ characterizes $h_2(m^\infty(1))$

Other characterizations by Thue (2)

m_2: $0 \mapsto 130402$, $1 \mapsto 132$, $2 \mapsto 1304$, $3 \mapsto 1402$, $4 \mapsto 1404$

- $\{AA, 010, 020\}$ characterizes $h_1(m_2^\infty(1))$
- $\{AA, 121, 212\}$ characterizes $h_2(m_2^\infty(1))$

$m_2^\infty(1)$ appears in both. Interesting ?

- Take a ternary word avoiding $\{AA, 010, 020\}$, delete each letter after a 0, you get a word avoiding $\{AA, 121, 212\}$.
- $\{AA, 01, 03, 10, 12, 20, 23, 24, 31, 34, 42, 43, 141, 302, 414, 2132, 3213\}$ characterizes $m_2^\infty(1)$.

Characterizations using $m_1^\infty(0)$

$$
\begin{aligned}
h_3\colon 0 &\mapsto 001011011101101001,\\
1 &\mapsto 00101101101001,\\
2 &\mapsto 00010.
\end{aligned}
$$

$$
\begin{aligned}
h_4\colon 0 &\mapsto 1001001101011001101001011001001101\\
&\quad\; 10010110100110110010011010010110011010 1,\\
1 &\mapsto 10010011010010 1,\\
2 &\mapsto 10010011011001011010011 01.
\end{aligned}
$$

- $\{AABBCABBA, 0011, 1100\}$ characterizes $h_3(m_1^\infty(0))$.
- Containing only 10101 as a 2^+-power and 11 distinct squares characterizes $h_4(m_1^\infty(0))$.

More examples by Golnaz Badkobeh

m_3: $0 \mapsto 032$, $1 \mapsto 04$, $2 \mapsto 12$, $3 \mapsto 14$, $4 \mapsto 1432$.

h_5: $0 \mapsto 1100$, $1 \mapsto 110$, $2 \mapsto 100$, $3 \mapsto 10$, $4 \mapsto 101100$.

h_6 : $0 \mapsto 100110$, $1 \mapsto 10010110$, $2 \mapsto 0110$, $3 \mapsto 01$,
$4 \mapsto 0100110$.

- Containing only 01010 and 10101 as 2^+-powers and 8 distinct squares characterizes $h_5(m_3^\infty(0))$.
- Containing only 1001001 as a 2^+-power and 14 distinct squares characterizes $h_6(m_3^\infty(0))$.

A third example with m_3

h_7: $0 \mapsto 00111001011000 1101$,

$1 \mapsto 0011100101100011010001011 0$,

$2 \mapsto 00111001011101$,

$3 \mapsto 0011100101110100111000110100010110$,

$4 \mapsto 001110010111010011100011010001011000 1101$.

$\{AABBCC, ABCABC, 0000, 1111, 0001110010110,$
$1110001101001, 0110100111000, 1001011000111\}$
characterizes $h_7(m_3^\infty(0))$.

Remarks

- $m_1^\infty(0)$ appears in many characterizations and we understand why.
- $m_2^\infty(1)$ appears in two similar characterizations and we see the link.
- $m_3^\infty(0)$ appears in three characterizations, seemingly by chance.

We hope to:

- Discover new useful (pure) morphic words and find their properties.
- Find links between characterizations using a same pure morphic word.
- Use them to quickly find and prove new characterizations.

Dejean words VS Pure morphic words

To obtain the avoidability index of patterns:

- I used morphic images of Dejean words.
  - Proves exponential factor complexity.
  - Automated proofs.
- Julien Cassaigne used (pure) morphic words.
  - More theoretical insights.
  - Mandatory for patterns with polynomial factor complexity.

The only word avoiding $AB.AC.BA.BC.CA$ is the fixed point Ω of $0 \mapsto 01$, $1 \mapsto 21$, $2 \mapsto 03$, $3 \mapsto 23$, up to permutation of letters.

$\Longrightarrow \{AB.AC.BA.BC.CA, 02, \cdots\}$ characterizes Ω.

One last case by Golnaz Badkobeh

m_4: $0 \mapsto 0102$,
$1 \mapsto 1013$,
$2 \mapsto 40135$,
$3 \mapsto 51024$,
$4 \mapsto 1024$,
$5 \mapsto 0135$.

h_8: $0 \mapsto 10011$,
$1 \mapsto 01100$,
$2 \mapsto 01001$,
$3 \mapsto 10110$,
$4 \mapsto 0110$,
$5 \mapsto 1001$.

Containing only 0110110 and 1001001 as 2^+-powers and 12 distinct squares characterizes $h_8(m_4^\infty(0))$.

Proof technique

To prove that $S = \mathcal{P} \cup \mathcal{F}$ characterizes a morphic word $h(m^\infty(0))$, where m is $\Sigma_{k'} \to \Sigma_{k'}$, h is $\Sigma_{k'} \to \Sigma_k$, and $k' > k$:

- Prove that extendable words in S are h-images of words over $\Sigma_{k'}$.
- Find a set $S' = \mathcal{P} \cup \mathcal{F}'$ that seem to characterize $m^\infty(0)$ over $\Sigma_{k'}$.
- Prove that S' characterizes $m^\infty(0)$ over $\Sigma_{k'}$:
  - Prove that extendable words in S' are m-images of words over $\Sigma_{k'}$.
  - Prove that the pre-images by m of words in S' are also in S'.

Remarks

- Conjecture: If $\lambda(P) = k$, then there exists a finite set S of forbidden patterns and factors containing P that characterizes a morphic word over Σ_k.
- The converse (every morphic word has a characterization) does not seem to hold: consider the Fibonacci word.
- For all these characterizations, the set can be put in the form $\{AB\cdots XAB\cdots X\} \cup \mathcal{F}$. It would be nice to have characterizations with other patterns.
- Strategy "from above" to prove avoidability.

Thank you for your attention !

View publication stats