Enumeration and Decidable Properties of Automatic Sequences

Émilie Charlier[1] Narad Rampersad [2] Jeffrey Shallit[1]

[1]University of Waterloo

[2]Université de Liège

Numération
Liège, June 6, 2011

k-automatic words

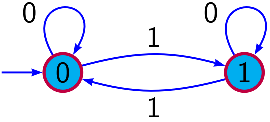
An infinite word $\mathbf{x} = (x_n)_{n \geq 0}$ is k-automatic if it is computable by a finite automaton taking as input the base-k representation of n, and having x_n as the output associated with the last state encountered.

Example

The Thue-Morse word is 2-automatic:

$$\mathbf{t} = t_0 t_1 t_2 \cdots = 011010011001 \cdots$$

It is defined by $t_n = 0$ if the binary representation of n has an even number of 1's and $t_n = 1$ otherwise.



Properties of the Thue-Morse word

- aperiodic
- uniformly recurrent
- contains no block of the form xxx
- contains at most $4n$ blocks of length $n+1$ for $n \geq 1$
- etc.

Enumeration and decidable properties

We present algorithms to decide if a k-automatic word

- is aperiodic
- is recurrent
- avoids repetitions
- etc.

We also describe algorithms to calculate its

- complexity function
- recurrence function
- etc.

Connection with logic

Theorem (Allouche-Rampersad-Shallit 2009)

Many properties are decidable for k-automatic words.

These properties are decidable because they are expressible as predicates in the first-order structure $\langle \mathbb{N}, +, V_k \rangle$, where $V_k(n)$ is the largest power of k dividing n.

Main idea

If we can express a property of a k-automatic word $\mathbf{x}$ using quantifiers, logical operations, integer variables, the operations of addition, subtraction, indexing into $\mathbf{x}$, and comparison of integers or elements of $\mathbf{x}$, then this property is decidable.

Another definition for k-automatic words

An infinite word $\mathbf{x} = (x_n)_{n \geq 0}$ is k-definable if,
for each letter a, there exists a FO formula φ_a of $\langle \mathbb{N}, +, V_k \rangle$ s.t.

$$\varphi_a(n) \text{ is true if and only if } x_n = a.$$

Theorem (Büchi-Bruyère)

An infinite word is k-automatic iff it is k-definable.

First direction: formula $\varphi \rightarrow$ DFA $\mathcal{A}_\varphi$

Second direction: DFA $\mathcal{A}_\varphi \rightarrow$ formula $\varphi_\mathcal{A}$

First direction: formula $\varphi \to$ DFA $\mathcal{A}_\varphi$

Automata for addition, equality and V_k are built in a straightforward way.

The connectives "or" and negation are also easy to represent.

Nondeterminism can be used to implement "$\exists$".

Ultimately, deciding the property we are interested in corresponds to verifying that $L(M) = \emptyset$ or that $L(M)$ is finite for the DFA M we construct.

Both can easily be done by the standard methods for automata.

Corollary (Bruyère 1985)

$\mathrm{Th}(\langle \mathbb{N}, + \rangle)$ *and* $\mathrm{Th}(\langle \mathbb{N}, +, V_k \rangle)$ *are decidable theories.*

Determining periodicity

Theorem (Honkala 1986)

Given a DFAO, it is decidable if the infinite word it generates is ultimately periodic.

It is sufficient to give the proof for k-automatic sets $X \subseteq \mathbb{N}$.

Let $\varphi_X(n)$ be a formula of $\langle \mathbb{N}, +, V_k \rangle$ defining X.

The set X is ultimately periodic iff

$$(\exists i)(\exists p)(\forall n)((n > i \text{ and } \varphi_X(n)) \Rightarrow \varphi_X(n+p)).$$

As $\mathrm{Th}(\langle \mathbb{N}, +, V_k \rangle)$ is a decidable theory, it is decidable whether this sentence is true, i.e., whether X is ultimately periodic.

Bordered factors

A finite word w is bordered if it begins and ends with the same word x with $0 < |x| \leq \frac{|w|}{2}$. Otherwise it is unbordered.

Example

The English word ingoing is bordered.

Theorem (C-Rampersad-Shallit 2011)

Let $\mathbf{x}$ *be a* k*-automatic word. Then the infinite word* $\mathbf{y} = y_0 y_1 y_2 \cdots$ *defined by*

$$y_n = \begin{cases} 1, & \textit{if } \mathbf{x} \textit{ has an unbordered factor of length } n; \\ 0, & \textit{otherwise}; \end{cases}$$

is k*-automatic.*

Arbitrarily large unbordered factors

Theorem (C-Rampersad-Shallit 2011)

The following question is decidable: given a k-automatic word $\mathbf{x}$, does $\mathbf{x}$ contain arbitrarily large unbordered factors.

Recurrence

An infinite word $\mathbf{x} = (x_n)_{n \geq 0}$ is recurrent if every factor that occurs at least once in it occurs infinitely often.

Equivalently, for each occurrence of a factor there exists a later occurrence of that factor.

Equivalently, for all n and for all $r \geq 1$, there exists $m > n$ such that for all $j < r$, $x_{n+j} = x_{m+j}$.

Uniform recurrence

An infinite word is uniformly recurrent if every factor that occurs at least once occurs infinitely often with bounded gaps between consecutive occurrences.

Equivalently, for all $r \geq 1$, there exists $t \geq 1$ such that for all n, there exists m with $n < m < n + t$ such that for all $i < r$, $x_{n+i} = x_{m+i}$.

Deciding recurrence

We obtain another proof of the following result:

Theorem (Nicolas-Pritykin 2009)

There is an algorithm to decide if a k-automatic word is recurrent or uniformly recurrent.

Some more results

Theorem (C-Rampersad-Shallit 2011)

Let $\mathbf{x}$ *be a* k*-automatic word. Then the following infinite words are also* k*-automatic:*

(a) $b(i) = 1$ *if there is a square beginning at position* i; 0 *otherwise*

(b) $c(i) = 1$ *if there is an overlap beginning at position* i; 0 *otherwise*

(c) $d(i) = 1$ *if there is a palindrome beginning at position* i; 0 *otherwise*

Brown, Rampersad, Shallit, and Vasiga proved results (a)–(b) for the Thue-Morse word.

Enumeration results

The *k*-kernel of an infinite word $(x_n)_{n\geq 0}$ is the set

$$\{(x_{k^e n+c})_{n\geq 0} \colon e \geq 0,\ 0 \leq c < k^e\}.$$

Theorem (Eilenberg)

An infinite word is k-automatic iff its k-kernel is finite.

k-regular sequences

With this definition we can generalize the notion of k-automatic words to the class of sequences over infinite alphabets.

A sequence $(x_n)_{n\geq 0}$ over $\mathbb{Z}$ is k-regular if the $\mathbb{Z}$-module generated by the set

$$\{(x_{k^e n+c})_{n\geq 0} \colon e \geq 0,\ 0 \leq c < k^e\}$$

is finitely generated.

Examples

- Polynomials in n with coefficients in $\mathbb{N}$
- The sum $s_k(n)$ of the base-k digits of n.

Factor complexity

The following result generalizes slightly a result of Mossé (1996).

Carpi and D'Alonzo (2010) proved a slightly more general result.

Theorem (C-Rampersad-Shallit 2011)

Let $\mathbf{x}$ *be a* k*-automatic word. Let* y_n *be the number of (distinct) factors of length* n *in* $\mathbf{x}$*. Then* $(y_n)_{n \geq 0}$ *is a* k*-regular sequence.*

Palindrome complexity

The following result generalizes a result of Allouche, Baake, Cassaigne and Damanik (2003).

Carpi and D'Alonzo (2010) proved a slightly more general result.

Theorem (C-Rampersad-Shallit 2011)

Let $\mathbf{x}$ *be a* k*-automatic word. Let* z_n *be the number of (distinct) palindromes of length* n *in* $\mathbf{x}$*. Then* $(z_n)_{n\geq 0}$ *is a* k*-regular sequence.*

Some more enumeration results

Theorem (C-Rampersad-Shallit 2011)

Let $\mathbf{x}$ *and* $\mathbf{y}$ *be* k*-automatic words. Then the following are* k*-regular:*

(a) *the number of (distinct) square factors in* $\mathbf{x}$ *of length* n*;*

(b) *the number of squares in* $\mathbf{x}$ *beginning at (centered at, ending at) position* n*;*

(c) *the length of the longest square in* $\mathbf{x}$ *beginning at (centered at, ending at) position* n*;*

(d) *the number of palindromes in* $\mathbf{x}$ *beginning at (centered at, ending at) position* n*;*

(e) *the length of the longest palindrome in* $\mathbf{x}$ *beginning at (centered at, ending at) position* n*;*

Theorem (cont'd)

(f) *the length of the longest fractional power in* $\mathbf{x}$ *beginning at (ending at) position* n;

(g) *the number of (distinct) recurrent factors in* $\mathbf{x}$ *of length* n;

(h) *the number of factors of length* n *that occur in* $\mathbf{x}$ *but not in* $\mathbf{y}$.

(i) *the number of factors of length* n *that occur in both* $\mathbf{x}$ *and* $\mathbf{y}$.

Brown, Rampersad, Shallit, and Vasiga proved results (b)–(c) for the Thue-Morse word.

Positional numeration systems

A positional numeration system is an increasing sequence of integers $U = (U_n)_{n\geq 0}$ such that

- $U_0 = 1$
- $(U_{i+1}/U_i)_{i\geq 0}$ is bounded $\quad \to C_U = \sup_{i\geq 0}\lceil U_{i+1}/U_i \rceil$

It is linear if it satisfies a linear recurrence over $\mathbb{Z}$.

The greedy U-representation of a positive integer n is the unique word $(n)_U = c_{\ell-1}\cdots c_0$ over $\Sigma_U = \{0,\ldots,C_U-1\}$ satisfying

$$n = \sum_{i=0}^{\ell-1} c_i\, U_i,\ \ c_{\ell-1} \neq 0 \ \text{ and } \ \forall t \ \sum_{i=0}^{t} c_i U_i < U_{t+1}.$$

U-automatic words

An infinite word $\mathbf{x} = (x_n)_{n\geq 0}$ is U-automatic if it is computable by a finite automaton taking as input the U-representation of n, and having x_n as the output associated with the last state encountered.

Example

Let $F = (1, 2, 3, 5, 8, 13, \ldots)$ be the sequence of Fibonacci numbers. Greedy F-representations do not contain 11.
The Fibonacci word

$$0100101001001010010100100101001\cdots$$

generated by the morphism $0 \mapsto 01,\ 1 \mapsto 0$ is F-automatic.
The $(n+1)$-th letter is 1 exactly when the F-representation of n ends with a 1.

Pisot systems

A Pisot number is an algebraic integer > 1 such that all of its algebraic conjugates have absolute value < 1.

A Pisot system is a linear numeration system whose characteristic polynomial is the minimal polynomial of a Pisot number.

An equivalent logical formulation

Let $V_U(n)$ be the smallest term U_i occurring in $(n)_U$ with a nonzero coefficient.

An infinite word $\mathbf{x} = (x_n)_{n\geq 0}$ is U-definable if, for each letter a, there exists a FO formula φ_a of $\langle \mathbb{N}, +, V_U \rangle$ s.t.

$$\varphi_a(n) \text{ is true if and only if } x_n = a.$$

Theorem (Bruyère-Hansel 1997)

Let U be a Pisot system. A infinite word is U-automatic iff it is U-definable.

Passing to this more general setting

By virtue of these results, all of our previous reasoning applies to U-automatic sequences when U is a Pisot system.

Hence, there exist algorithms to decide periodicity, recurrence, etc. for sequences defined in such systems as well.

What we can't do so far

k-automatic words are also generated by uniform morphisms (with some possible recoding of the alphabet).

The general case consists of morphic sequences: those generated by possibly non-uniform morphisms (again with a final recoding of the alphabet).

Some partial results are known (typically for purely morphic sequences and for U-automatic words).

Finding decision procedures for periodicity, etc. in the general setting remains an open problem.