数理解析研究所講究録
第 1655 巻 2009 年 106-110

Remarks on extractable submonoids

Genjiro Tanaka, Yoshiyuki Kunimochi

Dept. of Computer Science, Shizuoka Institute of Science and Technology,

Fukuroi-shi, 437-8555 Japan.

Masashi Katsura

Dept. of Mathematic, Kyoto Sangyo University,

Kita-Ku, Kyoto, 603-5555 Japan.

Key words: code, extractable code, bifix code, uniform code, free submonoid, free monoid.

Abstract. This paper deals with extractable codes. The base of free submonoid of a free monoid is called a code. The code C with the property that $z, xzy \in C^*$ implies $xy \in C^*$ is called an extractable code. Such C is a bifix code, and for example, appears as a Petri net code or a group code. In this paper we examine basic properties of extractable codes.

1. INTRODUCTION

Let A be an alphabet, A^* the free monoid over A, and 1 the empty word. Let $A^+ = A^* - \{1\}$. A word $v \in A^*$ is a ***right factor***(resp. left factor) of a word $u \in A^*$ if there is a word $w \in A^*$ such that $u = wv$(resp. $u = vw$). If v is a right factor of u, we write $v <_s u$. Similarly, we write $v <_p u$ if v is a left factor of u. For a word $w \in A^*$ and a letter $x \in A$ we let $|w|_x$ denote the number of x in w. The length $|w|$ of w is the number of letters in w. Therefore $A^n = \{w \in A^* |\ |w| = n\}$, $n \geq 1$. $Alph(w)$ is the set of all letters occurring at least once in w.

A non-empty subset C of A^+ is said to be a ***code*** if for $x_1, ..., x_p, y_1, ..., y_q \in C$, $p, q \geq 1$,

$$x_1 \cdots x_p = y_1 \cdots y_q \implies p=q,\ x_1=y_1,\ \ldots,\ x_p=y_p.$$

A subset M of A^* is a ***submonoid*** of A^* if $M^2 \subseteq M$ and $1 \in M$. Every submonoid M of a free monoid has a unique minimal set of generators $C = (M - \{1\}) - (M - \{1\})^2$. C is called the ***base*** of M. A submonoid M is ***right unitary*** in A^* if for all $u, v \in A^*$,

$$u,\ uv \in M \Longrightarrow v \in M.$$

M is called ***left unitary*** in A^* if it satisfies the dual condition. A submonoid M is ***biunitary*** if it is both left and right unitary. Let M be a submonoid of a free monoid A^*, and C its base. If $CA^+ \cap C = \emptyset$, (resp. $A^+C \cap C = \emptyset$), then C is called a ***prefix*** (resp. ***suffix***) code over A. C is called a ***bifix*** code if it is a prefix and suffix code. A submonoid M of A^* is right unitary (resp. biunitary) if and only if its minimal set of generators is a prefix code (resp. bifix code) ([1, p.46],[3, p.108]).

Let C be a nonempty subset of A^*. If $|x| = |y|$ for all $x, y \in C$, then C is a bifix code. We call such a code a *uniform code.* The uniform code A^n, $n \geq 1$, is called a *full uniform code.*
Let M be a submonoid of A^*. If the condition $z, xzy \in M$ implies $xy \in M$, then M is called an *extractable submonoid* of A^*.

If a submonoid N is extractable, then $u, 1uv \in C^*$ implies $1v = v \in C^*$. Similarly $v, uv \in C^*$ implies $u \in C^*$. Consequently M is biunitary. Therefore its minimal set of generators C is a bifix code.

Definition 1. Let $C \subset A^*$ be a code. If C^* is extractable, then C is called an extractable code.

1. FUNDMENTAL PROPERTIES OF EXTRACTABLE CODES

Let $\mathcal{A} = (Q, A, \delta, 1, F)$ be an automaton, where Q, A, $\delta : Q \times A \to Q$, 1, and F, are the state set, the input set, the next-state function, the initial state, and the final set of $\mathcal{A}$, respectively (For basic concepts of automata, refer to [4] or [1]). If for any $(p, q) \in Q \times Q$ there exists some $w \in A^*$ such that $\delta(p, w) = q$, then $\mathcal{A}$ is called transitive. If $\mathcal{A}$ has a fixed point s, and if for every $p, q \in Q, p \neq s$, there exists $w \in A^*$ such that $\delta(p, w) = q$, then $\mathcal{A}$ is called 0−transitive. If an automaton $\mathcal{A}$ is either transitive or 0-transitive, then $\mathcal{A}$ is called a [0]−transitive automaton.
Let L be a subset of A^*. For each $x \in A^*$, we define the set of all right contexts of x with respect to L by

$$Cont_L^{(r)}(x) = \{w \in A^* \mid xw \in L\}.$$

The right principal congruence $P_L^{(r)}$ of L is defined by $(x,\, y) \in P_L^{(r)}$ if and only if $Cont_L^{(r)}(x) = Cont_L^{(r)}(y)$. Let $u \in A^*$. By $[u]_L$ we denote the $P_L^{(r)}$−class of u by $[u]_L$ or simply by $[u]$. That is

$$[u]_L = \{v \mid Cont_{C^*}^{(r)}(v) = Cont_{C^*}^{(r)}(u), v \in A^*\}.$$

We denote by $[w_\phi]$ the class of $P_{C^*}^{(r)}$ consisting of all words $w \in A^*$ such that $wA^* \cap C^* = \phi$.
Let C be a prefic code. We defined the automaton $\mathcal{A}(C^*) = (A^*/P_{C^*}^{(r)}, A, \delta, [1], \{[1]\})$, where $\delta([w], x) = [wx]$ for $[w] \in A^*/P_{C^*}^{(r)}$ and $x \in A^*$. Then the automaton $\mathcal{A}(C^*)$, is minimal and [0]−transitive.

Let $\mathcal{A} = (Q, A, \delta, 1, \{1\})$ be a [0]−transitive minimal automaton recognizing C^* for some prefix code C. For each $p \in Q$ we put

$$W_p = \{w \in A^* \mid \delta(p, w) = 1\}.$$

Define the congruence ρ on $\mathcal{A}$ is by

$$p, q \in Q,\ p\rho q \Longleftrightarrow W_p = W_q.$$

Then ρ is the equality(See [2, p.215]).

Proposition 1. Let C be a prefix code, and let $\mathcal{A} = (Q, A, \delta, 1, \{1\})$ be a $[0]-$transitive automaton recognizing C^*. The following conditions are equivalent:
(1) C is extractable.
(2) $W_{\delta(p,z)} \subset W_p$ for all $p \in Q$, $z \in C^*$.

Corollary 2. Let C be a prefix code, and let A$= (Q, A, \delta, 1, \{1\})$ be a $[0]-$transitive minimal automaton recognizing C^*. If there exists some z_1, $z_2 \in C^*$ such that $\delta(p, z_1) = q$ and $\delta(q, z_2) = p$ for some $p, q \in Q$, $p \neq q$, then C is not extractable.

Let C be a prefix code, and let $\mathcal{A}(C^*) = (A^*/P_{C^*}^{(r)}, A, \delta, [1], \{[1]\})$ be the minimal automaton of C^*. Then, for $[x] \in A^*/P_{C^*}^{(r)}$ we have

$$u \in W_{[x]} \Leftrightarrow \delta([x], u) = [1] \Leftrightarrow xu \in C^* \Leftrightarrow u \in Cont_{C^*}^{(r)}(x).$$

That is, $W_{[x]} = Cont_{C^*}^{(r)}(x)$. Therefore we have the following rewriting of Proposition 1.

Proposition 3. Let C be a prefix code. Then the following two condition are equivalent:
(1) C is extractable.
(2) $Cont_{C^*}^{(r)}(xz) \subset Cont_{C^*}^{(r)}(x)$ for all $x \in A^*$ and $z \in C$.

Corollary 4. Let $C \subset A^*$ be a prefix code. If there exists some $z_1, z_2 \in C^*$ and some $[x], [y] \in A^*/P_{C^*}^{(r)}$, $[x] \neq [y]$, such that $[xz_1] = [y]$ and $[yz_2] = [x]$, then C^* is not extractable.

Corollary 5. Let $C \subset A^*$ be a prefix code. If either $[xz] = [x]$ or $[xz] = [w_\phi]$ for all $[x] \in A^*/P_{C^*}^{(r)}$ and $z \in C^*$, then C is extractable.

Proposition 6. Let $C \subset A^*$ be a prefix code such that $Cont_{C^*}^{(r)}(x) \cap Cont_{C^*}^{(r)}(y) = \phi$ for any distinct $[x], [y] \in A^*/P_{C^*}^{(r)}$. Then the following two conditions are equivalent
(1) C^* is an extractable code.
(2) Either $[xz] = [x]$ or $[xz] = [w_\phi]$ for any $x \in A^*$ and $z \in C$.

2. EXTRACTABLE REFLECTIVE CODE

Two words x, y are called conjugate if there exists words u, v such that $x = uv$, $y = vu$. the conjugacy relation is an equivalence relation. By $cl(x)$ we denote the class of x of this equivalence relation. Let $C \subset A^*$ be a code. If $uv \in C$ implies $vu \in C$, then C is called *reflective*. It is obvious that a reflective C is a union of conjugacy classes of words. Note that C^* is not necessarily a reflective language.
A code $C \subset A^+$ is called *infix* if for all

$$x, y, z \in A, \ \ xzy \in C \Longrightarrow x = y = 1.$$

Proposition 7. The reflective code C is an infix code.

Example 1. (1) The set $C = \{ab^2, bab, b^2a, a^3b, a^2ba, aba^2, ba^3\}$ is a union of two coujugacy classes of words in $\{a, b\}^*$. Since C is a prefix set, C is an infix code.
(2) Let $B \subset A$, $B \neq \emptyset$, and $n, k\,(k \leq n)$ be a positive integer. We let $|w|_B$ denote the number of letters of w which are in B. Then $U = \{w \in A^n \,|\, |w|_B = k\}$ is an extractable code. Since $uv \in U$ imples $vu \in U$, U is reflective.

Proposition 8. Let $C \subset A^*$ be a reflective code. The following two conditions are equivalent.
(1) C is extractable.
(2) For any $[x], [y] \in A^*/P_{C^*}^{(r)}$,

$$[x] \neq [y] \implies Cont_{C^*}^{(r)}(x) \cap Cont_{C^*}^{(r)}(y) = \phi.$$

Let $L \subset A^*$ and $n \geq 1$. We set

$$L^{(n)} = \{w^n \,|\, w \in L\}.$$

If D is a bifix code, then $D^{(n)}$ is a bifix code.

Proposition 9. If C is a reflective code, then $C^{(n)}, n \geq 1$, is a reflective code.

Proposition 10. Let D be a bifix code, and let $C = D^{(n)}$, $n \geq 2$. Then, for u, $v \in C(A^+)^{-1}$,

$$Cont_{C^*}^{(r)}(u) \cap Cont_{C^*}^{(r)}(v) \neq \phi \quad \implies \quad [u]_{C^*} = [v]_{C^*}.$$

From Proposition 10 and Proposition 8 we have

Corollary 11. If C is a reflective code, then $C^{(n)}, n \geq 2$, is an extractable reflective code.

3. EXTRACRTABLE UNIFORM CODES

In this section we examine extractable uniform codes.

Proposition 12. Let $C \subset A^*$ be a extractable code. Then $C \cap A^n$ is an extractable code.

Proposition 13. Let M be an extractable submonoid of A^*. Then $M \cap A^n$, $n \geq 1$, is an extractable code.

Example 2. Let S be a monoid and H an extractable submonoid. Let $\varphi : A^* \to S$ be a surjective morphism. Then $\varphi^{-1}(H)$ is an extractable submonoid of A^*.

Corollary 14. Let G be a group and H a normal subgroup of G. Let $\varphi : A^* \to G$ be a surjective morphism. Then $\varphi^{-1}(H) \cap A^n$ is an extractable reflective code.

Proposition 15. Let $C \subset A^+$ be a finite code with $A = alph(C)$. Then C is a maximal extractable code if and only if $C = A^n$ for some $n \geq 1$.

Proposition 16. Let $D \subset A^m$, $m \geq 1$, be a uniform code, then $D^{(n)}$, $n \geq 2$, is extractable.

Proposition 17. Let $w = (uv)^n u$, u, $v \in A^*$, $n \geq 2$, and let C be a conjugacy class of w.
(1) If $u = 1$ and $v \in A^+$, then C is extractable.
(2) If $u, v \in A^+$, then C is not extractable.

Proposition 18. Let $w \in A^+$, and let C be a conjugacy class of w.
(1) If $w = uv$, $u, v \in A^+$ and $Alph(u) \cap Alph(v) = \phi$. Then C is extractable.
(2) If $w = (uvu)^n(uv)^m, u, v \in A^+, m \geq 1, n \geq 0$ and $Alph(u) \cap Alph(v) = \phi$. Then C is not extractable.

Lemma 19. Let $x, y, u, v \in A^+$ with $|u|, |v| < |x| = |y|$. Then the following conditions hold.
(1) $x^2y = ux^2v \Rightarrow y = uv$ ($uv = x$ is not necessarily true).
(1') $xy^2 = uy^2v \Rightarrow x = uv$ ($uv = y$ is not necessarily true).
(2) $x^2y = uxyv \Rightarrow x = y = uv$ (uv is the power of some primitive word).
(2') $xy^2 = uyxv \Rightarrow x = y = uv$ (uv is the power of some primitive word).
(3) $x^2y = uyxv \Rightarrow x = y = uv$.
(3') $xy^2 = uyxv \Rightarrow x = y = uv$.
(4) $x^2y = uy^2v \Rightarrow y = uv$ (x and y are conjugate).
(4') $xy^2 = ux^2v \Rightarrow x = uv$ (x and y are conjugate).

Proposition 20. Let $x, y \in A^*$ with $|x| = |y| > 0$ and $C = \{x^2, xy, yx, y^2\}$. Then C is extractable.

References

[1] Berstel, J., and Perrin.D.: Theory of Codes. Academic Press, 1985
[2] Lallement, G.: Semigroup and Combinatorial Applications. Wiley. 1979.
[4] Tanaka. G.; "Limited codes associated with Petri nets", to be submitted.
[5] Tanaka. G., and Kunimochi. Y.: "Initial literal shuffles of uniform codes", to be submitted.