数理解析研究所講究録
第 786 巻 1992 年 100-106

SOME ALGEBRAIC PROPERTIES OF SEMICODES

H. J. Shyr and F. K. Tu

1 Introduction

Let X^* be the free monoid generated by a finite alphabet X, where $|X| \geq 2$. Every element of X^* is called a word and every subset of X^* is called a language. The empty word is denoted by 1 and let $X^+ = X^* \backslash \{1\}$. For any languages A and B the catenation of A and B is the set $AB = \{xy \mid x \in A, y \in B\}$ and A^+ is the set $A^+ = A \cup A^2 \cup A^3 \cup \cdots$. We let $A^{(n)} = \{u^n \mid u \in A\}$. Clearly, $A^{(n)} \subseteq A^n$. A language $A \subseteq X^+$ is a *code* if

$$x_1 x_2 \cdots x_n = y_1 y_2 \cdots y_m\,,\ x_i, y_j \in A,$$

$i = 1, 2, \cdots n$, $j = 1, 2, \cdots m$, implies $n = m$ and $x_i = y_i$ $i = 1, 2, \cdots, n$.

In this work we are interested in a biger class of languages which contains properly the class of all codes. The definition of a code A is that every word in X^+ has a unique representation over the set A. Our idea of semicode now is that every word in X^+ has more than two representation over A can happen only each side has equal number of words from A. Thus we give the following definition.

We call a language $A \subseteq X^+$ a *semicode* if

$$x_1 x_2 \cdots x_n = y_1 y_2 \cdots y_m\,,\ x_i, y_j \in A,$$

$i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, m$, implies $n = m$.

By the definition it is clear that every code is a semicode. It is immediate that a non-empty subset of a semicode is a semicode. We will see later that there are many semicodes which are not codes. The purpose of this work is a study of some algebraic properties of semicodes.

This is the abstract and the detail will be published eleswhere

2 Elementary algebraic properties of semicodes

First we present of the following:

Proposition 1 Let $A \subseteq X^+$. Then A is a semicode if and only if $A^p \cap A^q = \emptyset$ for all $p, q \geq 1, p \neq q$. ∎

Proposition 2 Let $A \subseteq X^+$. Then A is a semicode if and only if A^n is a semicode for every $n \geq 2$. ∎

We are now able to satate the following:

Proposition 3 Let $A \subseteq X^+$. Then the following are equivalent:

(1) A is a semicode;

(2) $A^p \cap A^q = \emptyset$, for all $p \neq q$, $p, q \geq 1$.;

(3) For some $n \geq 2$, A^n is a semicode;

(4) For every $n \geq 2$, A^n is a semicode;

(5) $A^+ = A \cup A^2 \cup A^3 \cup \cdots$ is a disjoint union. ∎

Since $A^{(n)}$ is a subset of A^n and a subset of a semicode is a semicode, we have

Corollary 4 If A is a semicode, then for every $n \geq 2$, $A^{(n)}$ is a semicode. ∎

Consider the language $A = \{ab, ab^2aba, a^2b, bab\}$. Clearly, $A^2 \cap A = \emptyset$ and $A^3 \cap A = \emptyset$. Since

$$(ab^2aba)(ab) = (ab)(bab)(a^2b)$$

The language A is not a semicode. This indicate that conditions $A^2 \cap A = \emptyset$ and $A^3 \cap A = \emptyset$ may not imply that A is a semicode.

There exists a language A such that $A^{(n)}$ is a semicode for every $n \geq 2$ but A is not a semicode. The following is an example. Let $A = \{a, ab, b\}$. Clearly A is a not semicode, while for every $n \geq 2$, $A^{(n)}$ is a semicode.

If A is a code and $n \geq 2$, then A^n is a code. Thus for any code A, A^n is a semicode for every $n \geq 2$.

For any language $A \subseteq X^+$, let $\sqrt{A} = \{f | f \in Q, f^n \in A \text{ for some } n \geq 1\}$. The language A a semicode may not implies $\sqrt{A}$ a semicode. The following is an example.

The language $A = \{a^2, aba, b^2\}$ is a semicode, while $\sqrt{A} = \{a, aba, b\}$ is not.

For the two element language we have

Proposition 5 Let $u \neq v \in X^+$. Then $\{u,v\}$ is a semicode if and only if $\{u,v\}$ is a code. ∎

Proposition 6 Let $A \subseteq X^+$ be semicode. Then $|A \cap f^+| \leq 1$ for all $f \in Q$. ∎

By a 2-code we shall mean a language $A \subseteq X^+$ such that every two elements from A forms a code.

Corollary 7 Every semicode A is a 2-code. ∎

The relation $\leq_c$ define on X^* as $u \leq_c v \iff u = xv = vx$ for some $x \in X^*$ is a partial order. By an independent set we shall mean that every two elements from the set is incomparable with respect the order $\leq_c$. $\leq_c$ indenpendent sets are the 2 codes. (see [3]).

Corollary 8 Let $A \subseteq X^+$ be a semicode. Then A is an independent set with respect to the order $\leq_c$. ∎

The above corollary tells us that the family of all semicodes is a subfamily of all $\leq_c$-independent sets (or all anti-commutative languages).

A word $w \in X^+$ is said to be *balanced* if there is a positive integer $k \geq 1$ such that $w_a = k$ for all $a \in X$. Here w_a means the number of letter a occures in the word w (see [3]).

Let $H = \{w \in X^* |\ w \text{ is balanced }\}$. Any subset of H is called a *balanced language.*

Proposition 9 If A is a balanced language and $A \subseteq X^m$ for some $m \geq 2$, then for any $u \notin H$ (i.e., u is not balanced), the language $A \cup uA$ is a semicode. ∎

Since every nonempty subset of a semicode is a semicode, we have the following corollary:

Corollary 10 If A is a balanced language and $A \subseteq X^m$ for some $m \geq 2$, then for $u \notin H$, the set $A \cup uA \cup Au$ is a semicode. ∎

Corollary 11 If A is a balanced language and $A \subseteq X^m$ for some $m \geq 2$, $u \notin H$, then $A \cup \{u\}$ is a semicode. ∎

It has been defined (see [3]) the partial orders $\leq_p$, $\leq_s$ on X^* as follows:

for $u, v \in X^*$, $u \leq_p v$ if $v = ux$ for some $x \in X^*$.

and

for $u, v \in X^*$, $u \leq_s v$ if $v = xu$ for some $x \in X^*$.

A language $A \subseteq X^*$ is $\leq_p$-independent if the following condition hold:
for any $u, v \in A$, $u \leq_p v$ implies that $u = v$.

A language $A \subseteq X^*$ is $\leq_s$-independent if the following condition hold:
for any $u, v \in A$, $u \leq_s v$ implies that $u = v$.

Now for a language $A \subseteq X^+$ we define $\underline{A}$ and $\overset{A}{\sim}$ to be the sets

$$\underline{A} = \{v \in A \mid w \leq_p v \Rightarrow w = v \text{ for all } w \in A\};$$
$$\overset{A}{\sim} = \{v \in A \mid w \leq_s v \Rightarrow w = v \text{ for all } w \in A\}.$$

That is $\underline{A}$ is the set of all minimal elements with respect to the partial order $\leq_p$ in A.

For any language $A \subseteq X^+$, $\underline{A}$ is always a prefix code. Similarly $\overset{A}{\sim}$ is the set of all minimal elements with respect to the partial order $\leq_s$ in A and $\overset{A}{\sim}$ is always a suffix code. For any $v \in \underline{A}$, we let

$$\bar{v} = \{w \in A \mid v \leq_p w\}.$$

Also for any $v \in \overset{A}{\sim}$, we let

$$\tilde{v} = \{w \in A \mid v \leq_s w\}.$$

It is clear that for any $u \neq v \in \underline{A}$,

$$\bar{u} \cap \bar{v} = \emptyset \quad \text{and} \quad A = \bigcup_{v \in \underline{A}} \bar{v}.$$

Also it is clear that for any $u \neq v \in \overset{A}{\sim}$,

$$\tilde{u} \cap \tilde{v} = \emptyset \quad \text{and} \quad A = \bigcup_{v \in \overset{A}{\sim}} \tilde{v}.$$

For any language $A \subseteq X^+$, by $A^{-1}A$ and AA^{-1} we shall mean the sets

$$A^{-1}A = \{x \in X^+ \mid w, wx \in A\};$$
$$AA^{-1} = \{x \in X^+ \mid w, xw \in A\}.$$

Remark : We note that if A is a semicode, then $A^{-1}A \cap A^p = \emptyset$ and $AA^{-1} \cap A^q = \emptyset$ for all q, $p \geq 1$. But the condition may not implies that A is a semicode as we can see from the following example.

Let $A = \{ab, abbaba, aab, bab\}$. Then $A^{-1}A = \{baba\}$ and $AA^{-1} = \{a, b\}$. Clearly $(A^{-1}A) \cap A^p = \emptyset$ and $(AA^{-1}) \cap A^q = \emptyset$ for all q, $p \geq 1$, while

$$(abbaba)(ab) = (ab)(bab)(aab)$$

and hence A is not a semicode.

Proposition 12 Let $A \subseteq X^+$. Then the following are equivalent:

(1) $fA \cap A = \emptyset$ for some $f \in A$

(2) $Af \cap A = \emptyset$ for some $f \in A$

(3) $A^{-1}A \cap A = \emptyset$.

(4) $AA^{-1} \cap A = \emptyset$. ∎

Proposition 13 Let $A, B \subseteq X^+$. Then the following are true:

(1) If $\underline{A} \cap \underline{B} = \emptyset$ and $\underline{A} \cup \underline{B}$ is a prefix code, then $fA^+ \cap A^+ = \emptyset$ for all $f \in B$.

(2) If $\underset{\sim}{A} \cap \underset{\sim}{B} = \emptyset$ and $\underset{\sim}{A} \cup \underset{\sim}{B}$ is a suffix code, then $A^+ f \cap A^+ = \emptyset$ for all $f \in B$. ∎

For a language $A \subseteq X^+$, define $\sigma(A)$ and $\sigma'(A)$ as following:

$\sigma(A) = \{f \in X^+ \,|\, A^n f \cap A \neq \emptyset$, for some $n \geq 1$ and $f \notin AX^*\}$;

$\sigma'(A) = \{f \in X^+ \,|\, fA^n \cap A \neq \emptyset$, for some $n \geq 1$ and $f \notin X^*A\}$.

It is clear that $A \cap \sigma(A) = \emptyset$ and $A \cap \sigma'(A) = \emptyset$.

Proposition 14 Let $A \subseteq X^+$. Then the following are equivalent:

(1) A is a code.

(2) $fA^+ \cap A^+ = \emptyset$ for all $f \in (A^{-1}A)$.

(3) $A^+ f \cap A^+ = \emptyset$ for all $f \in (AA^{-1})$.

(4) $fA^+ \cap A^+ = \emptyset$ for all $f \in \sigma(A)$.

(5) $A^+ f \cap A^+ = \emptyset$ for all $f \in \sigma'(A)$. ∎

Corollary 15 Let $A \subseteq X^+$ be a semicode. If A satified one of the following conditions, then A is a code.

(1) $\underline{A} \cup \underline{(A^{-1}A)}$ is a prefix code.

(2) $\underset{\sim}{A} \cup \underset{\sim}{(AA^{-1})}$ is a suffix code. ∎

Proposition 16 Let $A \subseteq X^+$ be a semicode. If A satified one of the following conditions, then A is a code.

(1) $\underline{A} \cup \underline{\sigma(A)}$ is a prefix code.

(2) $\underset{\sim}{A} \cap \underset{\sim}{\sigma'(A)}$ is a suffix code. ∎

Remarks:

1. The converse of the above proposition is not true. This can be seen from the following example.

Let $A = \{a, a^2\}$. Then $\sigma(A) = \sigma'(A) = \emptyset$ and we have $\underline{A} \cup \underline{\sigma(A)} = \{a\} = \underset{\sim}{A} \cap \underset{\sim}{\sigma'(A)}$ is a bifix code but A is not even a semicode.

2. If $\underline{A} \cup \underline{\sigma(A)}$ is not a prefix code, then there exists some $u, v \in \underline{A} \cup \underline{\sigma(A)}$ with $u <_p v$. By defintion of $\underline{\sigma(A)}$, we must have $v \in \underline{A}$ and $u \in \underline{\sigma(A)}$.

3. $A^{-1}A \not\subseteq \sigma(A)$ and $\sigma(A) \not\subseteq A^{-1}A$. Similarly $AA^{-1} \not\subseteq \sigma'(A)$ and $\sigma'(A) \not\subseteq AA^{-1}$.

For example : Let $A = \{ab, ba, abbabba\}$. Then we have $A^{-1}A = \{babba\}$, $\sigma(A) = \{bba\}$ and $AA^{-1} = \{abbab\}$, $\sigma'(A) = \{abb\}$.

Proposition 17 If A is a code, then the following are true:

(1) $(A^{-1}A) \cap (AA^{-1}) = \emptyset$.

(2) $\sigma(A) \cap \sigma'(A) = \emptyset$. ∎

Remark. The converse of the above proposition is not true. For example: Let $A = \{ab, abbaba, aab, bab\}$. Then $(A^{-1}A) = \{baba\}$, $(AA^{-1}) = \{a, b\}$ and we have $(A^{-1}A) \cap (AA^{-1}) = \emptyset$ but A is not even a semicode as we can see from the fact that $(abbaba)(bab) = (ab)(bab)(ab)(ab)$.

Proposition 18 Let $A \subseteq X^+$. Then the following are equivalent:

(1) $(A^{-1}A) \cap (AA^{-1}) = \emptyset$.

(2) $(A^{-1}A)A \cap A = \emptyset$.

(3) $A(AA^{-1}) \cap A = \emptyset$. ∎

Proposition 19 If $A \subset X^+$ is a semicode, then the following are true:

(1) $(A^{-1}A)A \cap A^n = \emptyset$ for all $n \geq 2$.

(2) $A(AA^{-1}) \cap A^n = \emptyset$ for all $n \geq 2$.

(3) $(A^{-1}A)A(AA^{-1}) \cap A^n = \emptyset$ for all $n \geq 2$. ∎

Proposition 20 Let A be a semicode. If $A \cap A^n f \neq \emptyset$ for some $f \in X^+ \backslash A$, $n \geq 1$, then $fA^i \cap A^j = \emptyset$ for all $i, j \geq 1$ and $i \neq (n-1) + j$. ∎

3 Mappings preserving Semicodes

A mapping $h : X^* \to X^*$ is a *homomorphism* if $h(xy) = h(x)h(y)$, for all $x, y \in X^*$. h is said to be *injective* if $h(x) = h(y)$ implies $x = y$, i.e., h is one to one.

We say that the homomorphism preserve semicodes if for every semicode A, $h(A)$ is a semicode. We have the following characerization of semicode preserving homomorphisms.

Proposition 21 Let $h : X^* \to X^*$ be a homomorphism. Then the following are equivalent:

(1) h is injective.

(2) h is code preserving.

(3) h is semicode preserving. ∎

REFERENCE

[1] M. Lothaire, Combinatorics on Words, Addison-Wesley Publi. co., Reading, Mass., 1983.

[2] R. C. Lyndon and M. P. Schützenberger, *The Equation* $a^M = b^N c^P$ *in a Free Group*, Michgan Math. J. **9** (1962) 289-298.

[3] H. J. Shyr, *Free monoids and languages.* Lecture Notes, Institute of Applied Mathematics, National Chung-Hsing University, Taichung, Taiwan, 1991

Institute of Applied Mathematics
National Chung-Hsing University
Taichung, Taiwan