数理解析研究所講究録
第 786 巻 1992 年 107-109

THE DECIPHERABILITY OF INFINITE CODED MESSAGES

S. S. Yu

Let the *alphabet* X be a finite nonempty set and let X^* be the *free monoid* generated by X. Elements of X^* are called *words* and subsets of X^* are called *languages*. We denote by λ the *empty word*, and by $X^+ = X^* \setminus \{\lambda\}$ the *free semigroup* generated by X. The *mutiplication* of two words x and y is defined as the juxtaposition of x and y, that is, the *catenation*. For any two languages A and B, the *catenation* of A and B is the set $AB = \{xy \mid x \in A,\ y \in B\}$. For a language L, let $L^n = LL^{n-1}$ and let $L^0 = \{\lambda\}$.

For a language $L \subseteq X^*$ and $x \in X^*$, an *L-factorization* of x is a sequence of words $(x_1, x_2, \ldots, x_n)$ in L such that $x = x_1 x_2 \cdots x_n$. For a nonempty language $L \subseteq X^+$, L is a *code* if every $x \in L^+$ has a unique L-factorization. For this reason, L is also called a uniquely decipherabile language. Elements of a code are called *codewords*. Any sequence of codewords is a *message*, i.e., $x \in L^+$ is a message from code L.

The decipherability of code has been investigated in three directions: the decoding of a finite message, an infinite message with known starting point, and an infinite message without the knowledge of its beginning and end.

For a finite message, the unique decipherability of a coded message is guaranteed by the definition of codes. But for an infinite message, the request for decipherability may not be the same. For example: Let $X = \{a, b\}$ and let $L = \{a, abb, bbab, abab\}$. Then L is a code. But the message $abbababab\cdots$ could be $(a)(bbab)(abab)\cdots$ or $(abb)(abab)(abab)\cdots$. We can not make sure which one is the correct meaning unless we know the end of this message.

We now give the following definition of the special decipherability of codes.

DEFINITION. [Ca] A code C has *finite decipherability delay* p for an integer $p \geq 0$ if and only if the equation $wxu = vy$ implies $w = v$ for all $w, v \in C$, $x \in C^p$, $u \in X^*$, and $y \in C^*$.

This is the abstract and the detail will be published eleswhere.

A *prefix code* $L \subseteq X^+$ is defined by $L \cap LX^+ = \emptyset$. Then every prefix code is a code having decipherability delay 0. Consider the following example: Let $X = \{a, b\}$ and let $L = \{ba, aa, bb\}$. Then L is a prefix code. But the message $\cdots ba \cdots$ could be $\cdots (ba) \cdots$ or $\cdots (bb)(aa) \cdots$. Even though we find a codeword ba in this message, we cannot make sure that this is the correct meaning of this word in this message if we do not know the beginning and the end of this message.

We give another definition which provides the ability to decode a message without knowing the beginning and the end of this message.

DEFINITION. [Be] A code C is called *synchronously decipherable* if there is an integer $s \geq 1$ such that

$$(*) \qquad x \in C^s,\ u, v \in X^*,\ uxv \in C^* \Rightarrow u, v \in C^*.$$

The smallest integer s such that $(*)$ holds is the synchronous decoding delay of C. It is denoted by $\iota(C)$.

If a code is synchronously decipherable, then there exists an integer s such that every s consecutive codewords in an infinite message can be decoded correctly without knowing the beginning and the end of this message.

PROPOSITION 1. If a code C is synchronously decipherable, then C has finite decipherability delay $q = 0$.

DEFINITION. A nonempty set $L \subseteq X^+$ is an *intercode of index m* if and only if $L^{m+1} \cap X^+ L^m X^+ = \emptyset$.

For properties of intercodes, see [Sh] and [Yu]. We have the following characterization of synchronously decipherable codes:

PROPOSITION 2. A code C is synchronously decipherable if and only if C is an intercode. Moreover, the smallest index of an intercode C is exactly the synchronous decoding delay $\iota(C)$ of C.

For a word $x \in X^+$, we define the set $E(x)$ as the set of all the factors of x, that is, $E(x) = \{w \mid x = ywz \text{ for some } y, z \in X^*\}$. Given a language L and a word w, the *L-representation of w* is defined as $w = x_1 y_1 x_2 y_2 \cdots x_n y_n x_{n+1}$, where $y_j \in L$, $i = 1, 2, \ldots, n$, $E(x_i) \cap L = \emptyset$, $i = 1, 2, \ldots, n+1$. If $E(w) \cap L = \emptyset$ or $L = \emptyset$, then $w = x_1$.

DEFINITION. A non-empty language $L \subseteq X^+$ is a *solid code* if for every $w \in X^+$ there is a unique L-representation of w.

When an information is transmitted over a noisy channel, some noise could be inserted into this information. A solid code can be used in information transmission over a noisy channel to allow us to decode those correct parts of a distributed message correctly. For properties of solid codes, see [Jü].

For $x \in X^*$, we define the proper left factor set $P(x)$ and the proper right factor set $S(x)$ of x as $P(x) = \{y \in X^+ \mid x \in yX^+\}$ and $S(x) = \{y \in X^+ \mid x \in X^+y\}$. Then we have the following characterization of solid codes:

PROPOSITION 3. A language $L \subseteq X^+$ is a solid code if and only if every two words $u, v \in L$ satisfy the following conditions:

(a) $P(u) \cap S(v) = \emptyset$,

(b) If $u \neq v$ then $u \notin E(v)$ and $v \notin E(u)$.

REFERENCES

[Be] J. Berstel and D. Perrin, *Theory of Codes*, Academic Press., Inc., Orlando; Toronto, 1985.

[Ca] R. M. Capocelli, *Finite Decipherability of Weakly Prefix Codes*, Colloquia Mathematics Societaties János Bolyai, **42**, Algebra, Combinatories and Logic in Computer Science, Györ, Hungary, (1983), 175–184.

[Jü] H. Jürgensen and S. S. Yu, *Solid Codes*, EIK: Journal of Information Processing and Cybernetics, to appear.

[Sh] H. J. Shyr and S. S. Yu, *Intercodes and Some Related Properties*, Soochow Journal of Mathematics **16**, no.1, (1990), 95–107.

[Yu] S. S. Yu, *A Characterization of Intercodes*, International Journal of Computer Mathematics, **36** (1990), 39–45.

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