Coding Partitions

Fabio Burderi and Antonio Restivo

Dipartimento di Matematica e Applicazioni

Palermo

Varese – Luglio 17-18, 2006

Let A be a finite alphabet. Let A^* denote the free monoid generated by A, and let $A^+ = A^* \setminus \{\varepsilon\}$.

A *code* X over A is a subset of A^+. The words of X are called *code words*, the elements of X^* *messages*, Let A be a finite alphabet. Let A^* denote the free monoid generated by A, and let $A^+ = A^* \backslash \{\varepsilon\}$.

A *code* X over A is a subset of A^+. The words of X are called *code words*, the elements of X^* *messages*,

A code X is said to be *uniquely decipherable* (UD) if every message has an unique factorization into codewords, i.e. the equality

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

implies $n = m$ and $x_1 = y_1, \ldots, x_n = y_n$.

Let X be a code and let $P = \{X_1, X_2, \ldots\}$, be a partition of X i.e. :
$\bigcup_{i \geq 1} X_i = X$ and $X_i \cap X_j = \emptyset$, for $i \neq j$.

A P-$factorization$ of an element $w \in X^+$ is a factorization

$$w = z_1 z_2 \cdots z_t$$

where:

- $\forall i \;\; z_i \in X_k^+$, $\quad X_k \in P$
- if $t > 1$, $z_i \in X_k^+ \Rightarrow z_{i+1} \notin X_k^+$, for all $1 \leq i \leq t-1$.

Example 1 $X = \{11, 00, 000, 111\} = \{x_1, x_2, x_3, x_4\}$,

$P = \{X_1, X_2, \}$, $X_1 = \{11, 00\}$, $X_2 = \{000, 111\}$. Let

$w = 1100000111 \in X^+$,

Example 1 $X = \{11, 00, 000, 111\} = \{x_1, x_2, x_3, x_4\}$,

$P = \{X_1, X_2, \}$, $X_1 = \{11, 00\}$, $X_2 = \{000, 111\}$. Let

$w = 1100000111 \in X^+$, $w = x_1 x_2 x_3 x_4 = x_1 x_3 x_2 x_4$

$z_1 z_2 = (1100)(000111)$ and $u_1 u_2 u_3 u_4 = (11)(000)(00)(111)$
are $P - factorizzations$ of z.

Example 1 $X = \{11, 00, 000, 111\} = \{x_1, x_2, x_3, x_4\}$,

$P = \{X_1, X_2, \}$, $\quad X_1 = \{11, 00\}$, $\quad X_2 = \{000, 111\}$. Let

$w = 1100000111 \in X^+$, $\quad w = x_1 x_2 x_3 x_4 = x_1 x_3 x_2 x_4$

$z_1 z_2 = (1100)(000111)$ and $u_1 u_2 u_3 u_4 = (11)(000)(00)(111)$
are $P - factorizzations$ of z.

The partition P is called a *coding partition* if any element $w \in X^+$ has a *unique* P-$factorization$, i.e. if

$$w = z_1 z_2 \cdots z_s = u_1 u_2 \cdots u_t,$$

with $\quad z_1 z_2 \cdots z_s, \quad u_1 u_2 \cdots u_t \quad$ P-$factorizations$ of w, then:
$s = t$ and $z_i = u_i$ for $i = 1, \ldots, s$.

P is *concatenatively independent* if, for $i \neq j$, $X_i^+ \cap X_j^+ = \emptyset$.

P is *concatenatively independent* if, for $i \neq j$, $X_i^+ \cap X_j^+ = \emptyset$.

Example 2 $X = \{00, 11, 000, 111\}$, $P = \{X_1, X_2, \}$,

$X_1 = \{00, 000\}$, $X_2 = \{11, 111\}$.

P is a coding partition of X.

P is *concatenatively independent* if, for $i \neq j$, $X_i^+ \cap X_j^+ = \emptyset$.

Example 2 $X = \{00, 11, 000, 111\}$, $P = \{X_1, X_2, \}$,

$X_1 = \{00, 000\}$, $X_2 = \{11, 111\}$.

P is a coding partition of X.

Example 3 $X = \{0, 01, 10\}$.

The word $w = 010 \in X^+$ has two factorizations :
$w = (0)(10) = (01)(0)$.

P is *concatenatively independent* if, for $i \neq j$, $X_i^+ \cap X_j^+ = \emptyset$.

Example 2 $X = \{00, 11, 000, 111\}$, $P = \{X_1, X_2, \}$,

$X_1 = \{00, 000\}$, $X_2 = \{11, 111\}$.

P is a coding partition of X.

Example 3 $X = \{0, 01, 10\}$.

The word $w = 010 \in X^+$ has two factorizations :
$w = (0)(10) = (01)(0)$.

P is a coding partition of X if and only if P is the trivial partition:
$P = \{X\}$.

Let X be a code and let $x_1x_2\cdots x_s = y_1y_2\cdots y_t$ be two factorizations of $w \in X^+$.

We say that the relation $x_1x_2\cdots x_s = y_1y_2\cdots y_t$ is *prime* if $\forall\, i < s, \forall\, j < t \;\Rightarrow x_1x_2\cdots x_i \neq y_1y_2\cdots y_j$.

A relation $w = x_1x_2\cdots x_s = y_1y_2\cdots y_t$, can be univocally factorized into prime relations.

Let X be a code and let $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$ be two factorizations of $w \in X^+$.

We say that the relation $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$ is *prime* if $\forall\, i < s, \forall\, j < t \;\Rightarrow x_1 x_2 \cdots x_i \neq y_1 y_2 \cdots y_j$.

A relation $w = x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$, can be univocally factorized into prime relations.

Theorem 1 Let $P = \{X_1, X_2, \ldots\}$ be a partition of a code X. P is a coding partition of X iff for every prime relation $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$ there exists a set X_h in the partition , such that $\forall\, i \leq s, \forall\, j \leq t,\, x_i,\, y_j \in X_h$.

Theorem 2 The set of the coding partitions of a code X is a complete lattice.

Given a code X, we can then define the finest coding partition P of X. It is called the *characteristic* partition of X and it is denoted by $P(X)$.

Theorem 2 The set of the coding partitions of a code X is a complete lattice.

Given a code X, we can then define the finest coding partition P of X. It is called the *characteristic* partition of X and it is denoted by $P(X)$.

- X is a UD code if and only if $P(X)$ is the *discrete partition*.

Theorem 2 The set of the coding partitions of a code X is a complete lattice.

Given a code X, we can then define the finest coding partition P of X. It is called the *characteristic* partition of X and it is denoted by $P(X)$.

- X is a UD code if and only if $P(X)$ is the *discrete partition*.

Given a partition $P = \{X_1, X_2, \ldots\}$ of a code X, a subset $Y \subseteq X$ is a *cross-section* of P, if $|Y \cap X_i| = 1$, for $i \geq 1$.

Theorem 3 If P is a coding partition of X, then any cross-section Y of P is a UD code.

A code X is called *ambiguous* if it is not UD. It is called *totally ambiguous* if $|X| > 1$ and $P(X)$ is the trivial partition.

The code $X = \{01, 10, 1\}$ is totally ambiguous.

Theorem 4 Let $P(X) = \{X_1, X_2, \dots\}$ be the characteristic partition of a code X, if $|X_i| > 1$ for some $i \geq 1$, then X_i is a totally ambiguous code.

Theorem 5 Let X be a code such that all proper subsets Y of X are UD codes. Then either X is a UD code or it is totally ambiguous.

Let X_0 be the union of all classes $Z \in P(X)$ such that $|Z| = 1$.
X_0 is a UD code and is called the *unambiguous component* of X.
The *canonical partition* of X is:

$$P_C(X) = \{X_0, X_1, \dots\},$$

where $|X_i| > 1$, for $i \geq 1$.

The sets X_i, with $i \geq 1$ are called the *totally ambiguous components* of X.

Example 4 Consider the code $X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\} = \{00, 0010, 1000, 11, 1111, 010, 011\}$.

$U_1 = \{(10, \{1, 2\}), (11, \{4, 5\})\}$,

$U_2 = \{(\varepsilon, \{4, 5\}), (00, \{1, 2, 3\}), (11, \{4, 5\})\}$,

$U_3 = \{(\varepsilon, \{1, 2, 3\}), (\varepsilon, \{4, 5\}), (10, \{1, 2, 3\}), (11, \{4, 5\})\}$,

$U_4 = U_2$.

So we have: $R_0 = \{6, 7\}, R_1 = \{1, 2, 3\}$, $R_2 = \{4, 5\}$.
Then $X_0 = \{010, 011\}, X_1 = \{00, 0010, 1000\}$, $X_2 = \{11, 111\}$
and $P_C(X) = \{X_0, X_1, X_2\}$ is the canonical partition of X.

If a code X is infinite, one can have partitions having an infinite number of classes and, moreover, each class can contain infinitely many elements.

If a code X is infinite, one can have partitions having an infinite number of classes and, moreover, each class can contain infinitely many elements.

Consider a partition having a finite number of classes and such that each class is a *rational* set.

Theorem 6 Given a partition $P = \{X_1, X_2, \ldots, X_n\}$ such that X_i, for $i = 1, 2, \ldots, n$, is a rational set, it is decidable whether P is a coding partition.

If a code X is infinite, one can have partitions having an infinite number of classes and, moreover, each class can contain infinitely many elements.

Consider a partition having a finite number of classes and such that each class is a *rational* set.

Theorem 6 Given a partition $P = \{X_1, X_2, \ldots, X_n\}$ such that X_i, for $i = 1, 2, \ldots, n$, is a rational set, it is decidable whether P is a coding partition.

Given a partition $P = \{X_1, X_2, \ldots, X_n\}$ of a code$X$, let $B = \{b_1, b_2, \ldots, b_n\}$ be an alphabet and let $\beta : X \rightarrow B^+$ be the following map: $x \mapsto b_i^{|x|}$ if $x \in X_i$. We can extend β to a morphism from X^* to B^* iff P is a coding partition of X.

Theorem 7 If X is rational, the number of classes of $P_C(X)$ is finite and each class of $P_C(X)$ is a rational set.

Remark further that, if X is not rational, then $P_C(X)$ can have infinitely many classes, as shown by the following example.

Example 5
Consider the code $X = \{a^n b, \quad c^n a^n b, \quad a^n b c^n \mid n \geq 1\}$.

It is easy to verify that $P_C(X)$ contains infinitely many classes and that any class X_i is of the form $X_i = \{a^i b, \quad c^i a^i b, \quad a^i b c^i\}$.

Let $P = \{X_1, X_2, \ldots\}$ be a coding partition of a code X, a component $X_i \in P$ is said maximal if $\forall w \notin X_i^+$ the partition $P' = \{X_1, \ldots, X_i \cup \{w\}, \ldots\}$ is not a coding partition.

A coding partition P is said maximal if every component of P is maximal.

Let $P = \{X_1, X_2, \dots\}$ be a coding partition of a code X, a component $X_i \in P$ is said maximal if $\forall w \notin X_i^+$ the partition $P' = \{X_1, \dots, X_i \cup \{w\}, \dots\}$ is not a coding partition.

A coding partition P is said maximal if every component of P is maximal.

We observe that if P is a maximal coding partition of X, then $\forall w \in A^+$ the partition $P' = P \cup \{\{w\}\}$ is not a coding partition of $X \cup \{w\}$.

We recall that a code X is said complete if $F(X^*) = A^*$, where $F(X^*)$ is the set of factors of code words.

Theorem 8 Let P a coding partition of a code X, if P is maximal then X is complete.

Theorem 8 Let P a coding partition of a code X, if P is maximal then X is complete.

If $P = \{X\}$ is the trivial partition of X, P is maximal iff $X^+ = A^+$.

Theorem 8 Let P a coding partition of a code X, if P is maximal then X is complete.

If $P = \{X\}$ is the trivial partition of X, P is maximal iff $X^+ = A^+$.

Theorem 9 Let X be a regular code and let P a non trivial coding partition of X. If X is complete then P is maximal.

Theorem 8 Let P a coding partition of a code X, if P is maximal then X is complete.

If $P = \{X\}$ is the trivial partition of X, P is maximal iff $X^+ = A^+$.

Theorem 9 Let X be a regular code and let P a non trivial coding partition of X. If X is complete then P is maximal.

Theorem 10 Let $P = \{X_1, X_2, \ldots, X_n\}$ be a coding partition of a code X and let $\mathcal{V}$ be a variety of codes. If $X_i \in \mathcal{V}$, for $i = 1, 2, \ldots, n$, then $X \in \mathcal{V}$.