Théorie des codes

Silvain Rideau

Langage formels,
calculabilité complexité

8 janvier 2009

Définition d'un code

Definition

Un sous-ensemble X de A^* est un **code** sur A si :

$$\forall m, n \geq 1 \text{ et } \forall (x_i)_{i=1\ldots n}, (x'_i)_{i=1\ldots m} \in X$$

$$x_1 x_2 \cdots x_n = x'_1 x'_2 \cdots x'_m \Rightarrow n = m \text{ et } \forall i \in 1 \ldots n \; x_i = x'_i$$

X est dit **maximal** s'il n'est pas strictement inclus dans un autre code.

Morphisme de codage

Définition alternative d'un code :

Proposition

Soit B un alphabet, tout morphisme $\phi : B^* \to A^*$ qui induit une bijection de B sur X est injectif.

Proposition (réciproque)

S'il existe un morphisme injectif $\phi : B^* \to A^*$ tel que $X = \phi(B)$ alors X est un code.

Cette définition est moins utile mais elle décrit bien la notion intuitive de codage et surtout de décodage.

Codes préfixes

- Un langage L est dit préfixe si aucun mot de L n'est prefixe d'un autre mot de L.
- Tout langage préfixe est un code.

Exemple

- $A = \{a, b\}$
- $X_1 = \bigsqcup_{n \geq 0} a^n b A^n$

X_1 est prefixe, c'est donc un code sur A.

Algorithme de reconnaissance des codes

Proposition

- $X \subset A^*$
- $U_1 = X^{-1}X \backslash \{\epsilon\}$
- $\forall n \geq 1\ U_{n+1} = X^{-1}U_n \cup U_n^{-1}X$

$$X \text{est un code} \iff \forall n \geq 1\ \epsilon \notin U_n$$

Le calcul des U_n donne l'algorithme.

Proposition

Si X est rationnel l'ensemble de ses U_n est fini.

Cet algorithme termine dans le cas des langages rationnels.

Un petit exemple

Exemple

- $A = \{a, b\}$
- $X_2 = \{aa, ba, bb, baa, bba\}$.

$U_1 = \{a\} = U_2$ donc X_2 est un code.

Distribution et mesure

Definition (Distribution de Bernoulli)

$\pi : A^* \to \mathbb{R}_+$ un morphisme tel que :

$$\sum_{a \in A} \pi(a) = 1$$

Une distribution est dite **positive** si : $\forall a \in A \ \pi(a) > 0$

Definition (Mesure d'un langage)

On étend π à $\mathcal{P}(A^*)$:

$$\pi(L) = \sum_{l \in L} \pi(l)$$

On vérifie aisément qu'on a bien une mesure sur A^*.

Mesure d'un code

$$\pi(LM) = \pi(\bigcup_{l\in L}\bigcup_{m\in M} lm) \leq \sum_{l\in L}\sum_{m\in M} \pi(lm) = \pi(L)\pi(M)$$

$$\pi(L^*) \leq \sum_{n\geq 0} \pi(L^n) \leq \sum_{n\geq 0} \pi(L)^n$$

Proposition

Soit X un code alors :

$$\forall n \geq 1\ \pi(X^n) = \pi(X)^n \text{ et } \pi(X^*) = \sum_{n\geq 0} \pi(X)^n$$

Réciproquement, si π est positive, si $\pi(X) < \infty$ et $\forall n \geq 1\ \pi(X^n) = \pi(X)^n$ alors X est un code.

Et encore des propriétés

Proposition

X un code sur A.

- $\forall \pi \ \pi(X) \leq 1$
- $\exists \pi$ *positve* $\pi(X) = 1 \Rightarrow X$ *maximal*

Ceci donne un critère pour prouver qu'un code est maximal, propriété pas toujours évidente :

Exemple

X_1 est maximal. On pose $\pi(a) = p < 1$ et donc $\pi(b) = 1 - p$.

$$\pi(X_1) = \sum_{n \geq 0} \pi(a^n b A^n) = \sum_{n \geq 0} p^n (1-p) \pi(A^n) = (1-p) \sum_{n \geq 0} p^n = 1$$

Complétude

On notera $F(L) = (A^*)^{-1}L(A^*)^{-1}$ l'ensemble des facteurs de L.

Definition (Ensembles denses, complets et maigres)

Complétude

On notera $F(L) = (A^*)^{-1}L(A^*)^{-1}$ l'ensemble des facteurs de L.

Definition (Ensembles denses, complets et maigres)

- $L \subset A^*$ est dense dans A^* si $A^* = F(L)$.

Complétude

On notera $F(L) = (A^*)^{-1}L(A^*)^{-1}$ l'ensemble des facteurs de L.

Definition (Ensembles denses, complets et maigres)

- $L \subset A^*$ est dense dans A^* si $A^* = F(L)$.
- Si P n'est pas dense, on dit que P est maigre.

Complétude

On notera $F(L) = (A^*)^{-1}L(A^*)^{-1}$ l'ensemble des facteurs de L.

Definition (Ensembles denses, complets et maigres)

- $L \subset A^*$ est dense dans A^* si $A^* = F(L)$.
- Si P n'est pas dense, on dit que P est maigre.
- $L \subset A^*$ est complet dans A^* si L^* est dense.

Complétude

On notera $F(L) = (A^*)^{-1}L(A^*)^{-1}$ l'ensemble des facteurs de L.

Definition (Ensembles denses, complets et maigres)

- $L \subset A^*$ est dense dans A^* si $A^* = F(L)$.
- Si P n'est pas dense, on dit que P est maigre.
- $L \subset A^*$ est complet dans A^* si L^* est dense.

Par exemple, X_1 est dense :

$$\forall w \in A^* \text{ on considère } a^{|w|}bw \in X_1.$$

Il est donc aussi complet.

Toujours des propriétés

A l'aide de nombreux autres résultats intermédiaires, et fastidieux, on montre aussi :

Proposition

Tout code maximal est complet

En fait on peut lui rajouter un mot qui n'est pas facteur et qui est sans bords.

Proposition

Soient X un langage maigre et complet et π positive :

$$\pi(X) \geq 1$$

On mélange tout

En rassemblant tous les résultats précédents, on obtient :

Théorème

Soit X un code maigre sur un alphabet A. Les propositions suivantes sont équivalentes :

On mélange tout

En rassemblant tous les résultats précédents, on obtient :

Théorème

Soit X un code maigre sur un alphabet A. Les propositions suivantes sont équivalentes :

(i) *X est un code maximal.*

On mélange tout

En rassemblant tous les résultats précédents, on obtient :

Théorème

Soit X un code maigre sur un alphabet A. Les propositions suivantes sont équivalentes :

(i) *X est un code maximal.*

(ii) *$\exists\ \pi$ positive sur A^* telle que $\pi(X) = 1$.*

On mélange tout

En rassemblant tous les résultats précédents, on obtient :

Théorème

Soit X un code maigre sur un alphabet A. Les propositions suivantes sont équivalentes :

(i) *X est un code maximal.*

(ii) *$\exists\ \pi$ positive sur A^* telle que $\pi(X) = 1$.*

(iii) *$\forall\ \pi$ positive sur A^*, $\pi(X) = 1$.*

On mélange tout

En rassemblant tous les résultats précédents, on obtient :

Théorème

Soit X un code maigre sur un alphabet A. Les propositions suivantes sont équivalentes :

(i) *X est un code maximal.*

(ii) *$\exists\ \pi$ positive sur A^* telle que $\pi(X) = 1$.*

(iii) *$\forall\ \pi$ positive sur A^*, $\pi(X) = 1$.*

(iv) *X est un code complet*

Pour en revenir à nos deux exemples

- X_1 vérifie ces 4 propriétés sans être maigre.
- X_2 est maigre (car fini). Posons $\pi(a) = p$ où $0 < p < 1$

$$\pi(X_2) = p^2 + (1-p)p + (1-p)^2 + (1-p)p^2 + (1-p)^2 p = 1$$

- X_2 est donc complet et maximal, ce qui a priori n'était pas évident.