ACCQ204 Dec. 10 2015

Cours 1

Enseignant: Aslan Tchamkerten *Crédit: Toni Franceschelli*

1 Un peu d'histoire...

La Théorie du codage date des années '50. Claude Elwood Shannon (1916-2001) et Richard Hamming (1915-1998) en sont les pionniers.

Le premier, considéré comme le père de l'ère digitale, s'est intéressé principalement aux *limites fondamentales* de communication en terme de:

- stockage de données: limite ultime de compression.
- transmission de données: limite ultime de vitesse de transmission fiable de données.

De façon complémentaire, Hamming s'est intéressé aux *algorithmes* permettant au mieux de corriger et détecter des erreurs. Le papier de Shannon *A mathematical theory of communication* (1948) et celui de Hamming *Error detecting and error correcting codes* (1950) établirent les domaine de la théorie de l'information et le domaine du codage, respectivement. A noter que Hamming considére un modéle de communication quelque peu différent de celui de Shannon.

2 Codes correcteurs d'erreurs

Problème de Hamming, exemple:

- On veut stocker des bits sur un support magnétique.
- Les bits sur le support peuvent se corrompre mais très rarement (au pire 1 bit sur 63).

2.1 Une solution naïve

Une première solution naïve consiste à répéter chaque bit 3 fois. La taille du mot code est donc 3 fois plus grande que celle du message. Exemple : message $\Rightarrow$ 0100 ; mot code $\Rightarrow$ 000111000000.

Performances:

- Complexité de codage et décodage: linéaire en la taille du message
- Taux de codage $= \frac{Taille\ message}{Taille\ mot\ code} = \frac{1}{3}$

Ce codage protège d'une erreur. Pour le décodage, on utilise la règle de la majorité sur 3 bits consécutifs.

2.2 Solution 1 de Hamming

On découpe le message en blocs de 4 bits chacun.
On associe à chaque bloc m un mot code $m \cdot G$ où $m \in \{0,1\}^4$ et

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Propriété:
$\forall m_1 \neq m_2 \in \{0,1\}^4$, $m_1 \cdot G$ et $m_2 \cdot G$ diffèrent d'au moins 3 positions.

Taux: $\frac{4}{7}$

Décodage:
Soit $y \in \{0,1\}^7$ contenant au plus 1 erreur.
$y \cdot H$ donne l'index du bit corrompu de y avec

$$H = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

2.3 Solution 2 de Hamming

$\exists$ deux matrices $G \in \mathcal{M}_{57,63}$ et $H \in \mathcal{M}_{63,6}$ possédant les propriétés suivantes:

- $\forall m_1 \neq m_2 \in \{0,1\}^4$, $m_1.G$ et $m_2.G$ diffèrent d'au moins 3 positions;
- si y est un mot "corrompu" d'au plus 1 erreur alors $y \cdot H$ donne l'index du bit corrompu;

Taux: $\frac{57}{63} > \frac{4}{7}$. Aucun schéma qui corrige une erreur ne peut atteindre un taux supérieur à $\frac{57}{63}$, comme on le verra plus bas.

2.4 Notions de Hamming

2.4.1 Distance de Hamming

Soit Σ un ensemble fini appelé alphabet.
Soit Σ^n l'ensemble des mots de n lettres sur Σ.

On appelle distance de Hamming $\Delta(x, y)$, avec $x, y \in \Sigma$, le nombre de coordonnées où x et y diffèrent.
On note $\delta(x, y)$ la distance normalisée de Hamming: $\delta(x, y) \stackrel{\text{def}}{=} \frac{\Delta(x,y)}{n}$.

Fait: La distance de Hamming est une métrique.

1. $\Delta(x, y) \geq 0, \forall x, y \in \Sigma^n$
2. $\Delta(x, y) = \Delta(y, x)$
3. $\Delta(x, y) + \Delta(y, z) \geq \Delta(x, z)$

2.4.2 Codes

Soit $\mathcal{C} \subseteq \Sigma^n$.

1. $\mathcal{C}$ corrige t erreurs si tout motif de t erreurs peut être corrigé (par un décodage possiblement inefficace).
   Formellement:
   - $B(x, t) \stackrel{\text{def}}{=} \{y \in \Sigma^n : \Delta(x, y) \leq t\}$
   - $\mathcal{C}$ corrige t erreur si $\forall x, y \in \mathcal{C}$ avec $x \neq y$, $B(x, t) \cap B(y, t) = \emptyset$.

2. $\mathcal{C}$ détecte $e \geq 1$ erreurs si à chaque fois que $1 \leq \#\ erreurs \leq e$, on peut détecter que des erreurs ont eu lieu.
   Formellement:
   $\forall x \in \mathcal{C}, B(x,e) \cap C = \{x\}$

3. On appelle distance d'un code $\Delta(\mathcal{C})$, la distance minimale qui sépare deux mots d'un code:

$$\Delta(\mathcal{C}) = \min_{x,y \in \mathcal{C}, x \neq y} \Delta(x,y)$$

Proposition 1 *Les conditions suivantes sont équivalentes:*

1. $\mathcal{C}$ *corrige* t *erreurs*
2. $\mathcal{C}$ *détecte* $2t$ *erreurs*
3. $\Delta(\mathcal{C}) \geq 2t+1$

Preuve

- $3 \rightarrow 1$:
  $\Delta(\mathcal{C}) \geq 2t+1 \Rightarrow$ Les boules $B(x,t)$ ne se recouvrent pas $\Rightarrow$ on associe à $y \in \Sigma^n$ le décodage "plus proche voisin"

$$\Phi(y) = \arg\min_{x \in \mathcal{C}} \Delta(x,y)$$

  Ce décodeur corrige bien t erreurs.

- $\neg 3 \rightarrow \neg 1$:
  $\Delta(\mathcal{C}) \leq 2t \Rightarrow \exists$ 2 mots codes x_1 et $x_2 \in \mathcal{C}$ dont les boules de rayon t se recouvrent: $B(x_1,t) \cap B(x_2,t) \neq \emptyset$.
  Si y appartient à cette intersection $\rightarrow$ problème pour décoder.

- $3 \rightarrow 2$:
  $\forall x \in \mathcal{C}, B(x,2t) \cap C = \{x\}$
  On considère le décodage:
  Si $\left|\begin{array}{l} y^n = x^n \in \mathcal{C}, \text{ on déclare } x^n. \\ y^n \in \cup_x B(x,2t) \backslash \mathcal{C}, \text{ on déclare "erreur".} \\ y \notin \cup_x B(x,2t), \text{ on déclare n'importe quel mot code.} \end{array}\right.$
  Ce décodeur détecte bien $2t$ erreurs.

- $\neg 3 \rightarrow \neg 2$:
  $\Delta(\mathcal{C}) \leq 2t \Rightarrow$ 2 mots codes appartiennent à une même boule $\Rightarrow$. Si y est égal à l'un de ces mots codes il n'est pas possible de savoir si y correspond à un mot code où s'il s'agit d'une version bruitée d'un mot code.

■

Proposition 2 *Pour* $\Sigma = \{0, 1\}$

1. $|B(x,t)| = \sum_{i=0}^{t} \binom{n}{i} \stackrel{\text{def}}{=} \text{Vol}(n,t)$
2. *Si* $\mathcal{C}$ *corrige* t *erreurs* $\Rightarrow |\mathcal{C}| \leq \frac{2^n}{\text{Vol}(n,t)}$
3. *Soit* $0 \leq p \leq \frac{1}{2}$ *alors*
   (a) $\text{Vol}(n, np) \leq 2^{nH(p)}$ *pour tout* np *entier*
   (b) $\text{Vol}(n, np) \geq 2^{n(H(p)-o(1))}$ *pour* n *suffisament grand*

   où $H(p) \stackrel{\text{def}}{=} -p \cdot log_2(p) - \overline{p} \cdot log_2(\overline{p})$, $\overline{p} \stackrel{\text{def}}{=} 1 - p$

Observation 3 $H(p)$, $0 \leq p \leq 1$, *est une fonction concave, symétrique, qui atteint son maximum à* $p = 1/2$, *et telle que* $H(0) = H(1) = 0$.

Observation 4 *Pour* $n = 63$, $t = 1$ *on a* $\text{Vol}(63, 1) = 64 \Rightarrow |\mathcal{C}| \leq \frac{2^{63}}{64} = 2^{57}$ $\Rightarrow$ *Taux* $\frac{57}{63}$ *optimal (Solution 2 Hamming).*

Preuve

1. $\binom{n}{i}$ représente le nombre de séquences de longueur n qui diffèrent d'une séquence donnée sur i coordonnées exactement.
2. Si $\mathcal{C}$ corrige t erreurs alors pour tout $x, y \in \mathcal{C}$ on a $B(x,t) \cap B(y,t) = \emptyset$ d'où
$$2^n \geq |\cup_{x \in \mathcal{C}} B(x,t)| = |\mathcal{C}| \cdot \text{Vol}(n,t)$$

3. (a)

$$\begin{aligned}\sum_{i=0}^{np}\binom{n}{i} &= 2^{nH(p)}\sum_{i=0}^{np}\binom{n}{i}\cdot p^{np}\cdot\overline{p}^{n\bar{p}}\\ &= 2^{nH(p)}\sum_{i=0}^{np}\binom{n}{i}\cdot p^{i}\cdot\overline{p}^{n-i}\cdot\left(\frac{p}{1-p}\right)^{np-i}\\ &\leq 2^{nH(p)}\end{aligned}$$

où l'inégalité vient du fait que $p/(1-p) \leq 1$ pour $p \leq 1/2$ et de l'identité

$$\sum_{i=0}^{n}\binom{n}{i}\cdot p^{i}\cdot\overline{p}^{n-i} = 1.$$

(b) En utilisant une version grossière de la formule de Stirling

$$k! = k^{k}\cdot e^{-k}\text{poly}(k)$$

où poly(k) est un terme polynomiale en k (i.e., $k^{\alpha} \leq \text{poly}(k) \leq k^{\beta}$ pour certains $0 < \alpha \leq \beta$ et k suffisamment grand) on a

$$\begin{aligned}\sum_{i=0}^{np}\binom{n}{i} &\geq \binom{n}{np}\\ &= \left(\frac{1}{p}\right)^{pn}\left(\frac{1}{\bar{p}}\right)^{\bar{p}n}\text{poly}(n)\\ &= 2^{nH(p)}\text{poly}(n)\\ &\geq 2^{n(H(p)-o(1))}\end{aligned}$$

■