

誤り訂正符号

通信途中でのノイズ (雑音) による誤りに
どう対処するか

→ 多少の誤りなら検出・訂正できる仕組みを
数理技術により実現 (誤り訂正符号)

情報通信中に誤り (らしきこと) に出遭ったとき

- 聞き直す (より安全なプロトコルの採用)
- 見当を付ける (誤りの自動訂正)

→ 冗長性を利用して安全性を確保

誤り訂正符号 (お話)

「誤りの自動訂正」が出来るように
予め適切に冗長性を持たせて通信する

- 誤り訂正性能は高く
- とは言えなるべく効率的に

→ 有限体上の線型代数・代数幾何などの
数理論理構造を利用

誤り訂正符号

誤り検出:

符号 $C^* : S^* \rightarrow T^*$ で、
受信した語 $y \in T^*$ が $y \notin \text{Im}C^*$ なら誤り

誤り訂正:

正しくは y に “一番近い” $x \in \text{Im}C^*$ だろう

誤り訂正符号

受信語 $y \notin \text{Im}C^* \subset T^*$ で誤り検出

→ T^* 全部は使わない

→ 冗長度を持たせて誤り検出・訂正

とは言え

- より効率良く (冗長度少なめ)
- より高い誤り対処性能を持つ
(誤りが沢山あっても大丈夫)

ものが望ましい

誤り訂正符号

以下では、

- 生起確率の違いを考慮しない
- 等長符号のみを考える
(全ての符号語が同じ長さ)

効率良い情報源符号で符号化された文字列を
一定の個数毎に切って再符号化する
と想定 (通信路符号)

誤り訂正符号

符号

$$\mathcal{C} : S \longrightarrow V := T^n \quad (n : \text{符号語長})$$

の像 $\text{Im}\mathcal{C} =: U \subset V$ のみが大事

→ 寧ろ、像 U をも単に \mathcal{C} と書き、
これを符号と呼ぶ： $\mathcal{C} \subset V$

誤り訂正符号の性能を表すには

- 符号長 n (当座固定)
- 符号語数 $M := \#C$ (大きいほど良い)
- 訂正出来る誤りの数 t (大きいほど良い)

しかし一般に、

「 $M \rightarrow$ 大」と「 $t \rightarrow$ 大」とは

相反する要求!!

誤り訂正符号の性能を表すには

- 符号長 n (当座固定)
- 符号語数 $M := \#C$ (大きいほど良い)
- 訂正出来る誤りの数 t (大きいほど良い)

しかし一般に、

「 $M \longrightarrow$ 大」と「 $t \longrightarrow$ 大」とは

相反する要求!!

誤りの訂正

符号 $\mathcal{C} \subset V = T^n$ で、

- 受信語 $y \notin \mathcal{C}$ によって誤り検出
- y に “一番近い” $x \in \mathcal{C}$ が正しい、
として誤り訂正

→ “一番近い” とは？

→ V に “距離” を導入
(通常 Hamming 距離 を用いる)

誤りの訂正

符号 $\mathcal{C} \subset V = T^n$ で、

- 受信語 $y \notin \mathcal{C}$ によって誤り検出
- y に “一番近い” $x \in \mathcal{C}$ が正しい、
として誤り訂正

→ “一番近い” とは？

→ V に “距離” を導入
(通常 Hamming 距離 を用いる)

誤りの訂正

符号 $\mathcal{C} \subset V = T^n$ で、

- 受信語 $y \notin \mathcal{C}$ によって誤り検出
- y に “一番近い” $x \in \mathcal{C}$ が正しい、
として誤り訂正

→ “一番近い” とは？

→ V に “距離” を導入
(通常 **Hamming 距離** を用いる)

距離の公理

X : 集合

$d : X \times X \longrightarrow \mathbf{R}_{\geq 0}$: X 上の距離
(metric, distance)

であるとは、

- $d(x, y) = 0 \iff x = y$
- $d(x, y) = d(y, x)$
- $d(x, y) + d(y, z) \geq d(x, z)$
: 三角不等式 (triangle inequality)

Hamming 距離

$V = T^n$ 上に次で定まる距離

$$d : V \times V \longrightarrow \mathbf{R}_{\geq 0}$$

を V 上の **Hamming 距離 (distance)** と呼ぶ :

$\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_n) \in V$ に対し、

$$d(\mathbf{x}, \mathbf{y}) := \#\{i \mid x_i \neq y_i\}$$

誤り訂正の性能

$$\mathcal{C} \subset V = T^n$$

n 箇所のうち何箇所違ってても訂正できるか？

= 距離が幾ら以内なら訂正できるか？

t 箇所違ってても一意に訂正できる

$\iff \forall \mathbf{y} \in V$ に対し

$d(\mathbf{x}, \mathbf{y}) \leq t$ なる $\mathbf{x} \in \mathcal{C}$ は高々 1 つ
($\#\{\mathbf{x} \in \mathcal{C} \mid d(\mathbf{x}, \mathbf{y}) \leq t\} \leq 1$)

誤り訂正の性能

$$\mathcal{C} \subset V = T^n$$

n 箇所のうち何箇所違っても訂正できるか？

= 距離が幾ら以内なら訂正できるか？

t 箇所違っても一意に訂正できる

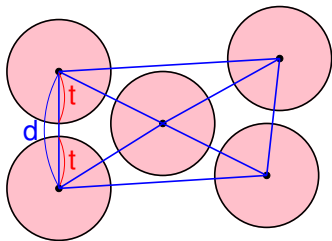
$\iff \forall \mathbf{y} \in V$ に対し

$$d(\mathbf{x}, \mathbf{y}) \leq t \text{ なる } \mathbf{x} \in \mathcal{C} \text{ は高々 1 つ}$$
$$(\#\{\mathbf{x} \in \mathcal{C} \mid d(\mathbf{x}, \mathbf{y}) \leq t\} \leq 1)$$

誤り訂正の性能

$d := \min\{d(\mathbf{x}, \mathbf{x}') \mid \mathbf{x}, \mathbf{x}' \in \mathcal{C}, \mathbf{x} \neq \mathbf{x}'\}$
: \mathcal{C} の**最小距離 (minimum distance)**

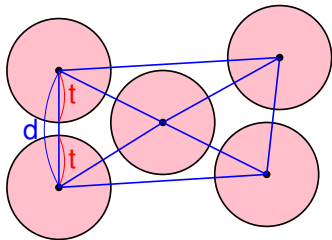
$d \geq 2t+1$ なら一意に訂正可能 $\longrightarrow t = \left\lfloor \frac{d-1}{2} \right\rfloor$



誤り訂正の性能

$d := \min\{d(\mathbf{x}, \mathbf{x}') \mid \mathbf{x}, \mathbf{x}' \in \mathcal{C}, \mathbf{x} \neq \mathbf{x}'\}$
: \mathcal{C} の最小距離 (**minimum distance**)

$d \geq 2t+1$ なら一意に訂正可能 $\longrightarrow t = \left\lfloor \frac{d-1}{2} \right\rfloor$



誤り訂正符号の性能を表すには

- 符号長 n (当座固定)
- 符号語数 $M := \#C$ (大きいほど良い)
- 訂正出来る誤りの数 t (大きいほど良い)

→ 訂正性能 t は最小距離 d で計れる!!

誤り訂正符号の性能を表すには

- 符号長 n (当座固定)
- 符号語数 $M := \#C$ (大きいほど良い)
- 最小距離 d (大きいほど良い)

→ (n, M, d) -符号

誤り訂正符号の性能を表すには

- 符号長 n (当座固定)
- 符号語数 $M := \#C$ (大きいほど良い)
- 最小距離 d (大きいほど良い)

「 $M \longrightarrow$ 大」と「 $d \longrightarrow$ 大」とは
相反する要求!!

問題: n, d を固定した時の M の上限は ?

符号語数の評価

問題: n, d を固定した時の M の上限は ?

d : 設計距離

$$A_q(n, d) := \max\{M \mid \exists \mathcal{C} : q \text{ 元 } (n, M, d)\text{-符号}\}$$

→ n, d と $q := \#T$ とで評価する

符号語数の評価

Hamming 距離で “半径 t の球” の元の個数は?

$B(\mathbf{x}, t) := \{\mathbf{y} \in V \mid d(\mathbf{x}, \mathbf{y}) \leq t\}$
: 中心 \mathbf{x} , 半径 t の球体

$$\#B(\mathbf{x}, t) = \sum_{k=0}^t \binom{n}{k} (q-1)^k?$$

符号語数の評価

Hamming 距離で “半径 t の球” の元の個数は?

$B(\mathbf{x}, t) := \{\mathbf{y} \in V \mid d(\mathbf{x}, \mathbf{y}) \leq t\}$
: 中心 \mathbf{x} , 半径 t の球体

$$\#B(\mathbf{x}, t) = \sum_{k=0}^t \binom{n}{k} (q-1)^k$$

Hamming の球充填上界

(sphere-packing bound)

$q := \#T$ のとき、

$(n, M, 2t + 1)$ -符号が存在

$$\implies M \left(\sum_{k=0}^t \binom{n}{k} (q-1)^k \right) \leq q^n$$

従って、

$$A_q(n, d) \left(\sum_{k=0}^t \binom{n}{k} (q-1)^k \right) \leq q^n \quad (t = \lfloor \frac{d-1}{2} \rfloor)$$

Gilbert-Varshamov の下界

$q := \#T$ のとき、

$$A_q(n, d) \left(\sum_{k=0}^{d-1} \binom{n}{k} (q-1)^k \right) \geq q^n$$

即ち、

$$M \left(\sum_{k=0}^{d-1} \binom{n}{k} (q-1)^k \right) \leq q^n$$

$\implies (n, M, d)$ -符号が存在

問題: n, d を固定した時の M の上限は ?

符号語数 M の大きい符号を
組織的に構成するには、

C の元がなるべく “均等に” 分布する
のが望ましい

→ $V = T^n$ の持つ数理構造 (対称性) を利用
... 線型符号