数理の世界・不思議体験 ２００３

「符号理論入門 —デジタルの数学—」

２００３年9月１３日 (土)

我々の身の回りには色々な形で数理科学が役に立っています. 携帯電話での会話, 音楽再生やコンピュータのデータ保存のためのコンパクト・ディスクなどのデータのやり取りにはデジタル通信が用いられています. このデジタルの世界に欠かせない数理科学の一つの分野が (誤り訂正) 符号理論と呼ばれるものです.

ある種の情報をデジタルの通信路を通して伝える際に, 通信路には雑音が発生し, 送信した情報を受信した際には情報に誤りが含まれる場合があり, 受信者は間違った情報を受け取ることになります. このとき, 誤り訂正符号を用いれば, 少ない間違いであれば自動的に誤りを訂正して正しい情報として受信することが可能になります.

この講義では, 符号理論の基礎として, ハミング符号 e_7 を用いてどのように誤りが訂正されるのか, なぜ, 誤りを訂正することが出来るのか, などを高校生の皆さんにも理解出来るように出来るだけ簡単に説明したいと思います.

山形大学 理学部 数理科学科 原田 昌晃

1　準備

- ハミング符号 e_7 には１６個のベクトルがあります. まずはそれをリストアップします.

$$\begin{array}{ll}
\overrightarrow{c_1} = (0,0,0,0,0,0,0) & \overrightarrow{c_9} = (1,0,1,1,0,1,0) \\
\overrightarrow{c_2} = (1,0,0,1,1,0,0) & \overrightarrow{c_{10}} = (0,1,1,0,0,1,1) \\
\overrightarrow{c_3} = (0,1,0,0,1,0,1) & \overrightarrow{c_{11}} = (1,1,0,0,1,1,0) \\
\overrightarrow{c_4} = (1,1,1,0,0,0,0) & \overrightarrow{c_{12}} = (1,0,1,0,1,0,1) \\
\overrightarrow{c_5} = (0,0,1,0,1,1,0) & \overrightarrow{c_{13}} = (0,1,1,1,1,0,0) \\
\overrightarrow{c_6} = (1,0,0,0,0,1,1) & \overrightarrow{c_{14}} = (1,1,0,1,0,0,1) \\
\overrightarrow{c_7} = (0,1,0,1,0,1,0) & \overrightarrow{c_{15}} = (0,0,0,1,1,1,1) \\
\overrightarrow{c_8} = (0,0,1,1,0,0,1) & \overrightarrow{c_{16}} = (1,1,1,1,1,1,1)
\end{array}$$

- [伝言ゲーム] ５人一組で次のような伝言ゲームをします.

  (1) e_7 のベクトルの中から一つ選ぶ.

  (2) 暗記して次の人に伝える.

  (3) 最後の人は伝えられたベクトルが e_7 の１６個のベクトルにあるかどうか調べる.

    (3-1) もしあればそれを最初に送ったベクトルとみなす.

    (3-2) ない場合は伝えられたベクトルの一ヶ所の 0 と 1 を入れ換えたものが１６個の中にあるかどうか調べる. あれば, それを最初に送ったベクトルとみなす.

    (3-3) それでもなければ, ゲームオーバー.

2　符号の定義

位数２の有限体：$\mathbb{F}_2 = \{0,1\}$

$+$	0	1
0	0	1
1	1	0

$\times$	0	1
0	0	0
1	0	1

$\mathbb{F}_2^n = \{(x_1, x_2, \ldots, x_n) | x_i \in \mathbb{F}_2\}$
0 と 1 の n 個の数字の組 (ベクトル) 全体

例 1. $\mathbb{F}_2^3 =$

$\mathbb{F}_2^n$ のベクトル $\overrightarrow{x} = (x_1, \ldots, x_n)$, $\overrightarrow{y} = (y_1, \ldots, y_n)$ に対して, その和を $\overrightarrow{x} + \overrightarrow{y} = (x_1 + y_1, \ldots, x_n + y_n)$ で定義する.

定義 2. $\mathbb{F}_2^n$ の部分集合 C において全ての C の元 (ベクトル) $\overrightarrow{x}, \overrightarrow{y}$ に対して $\overrightarrow{x} + \overrightarrow{y}$ が C の元であれば C を 符号 とよぶ.

ここで, 各符号 C には 2^k 個のベクトルが含まれることが知られている.

例 3. $C_1 = \{(0,0,0), (1,1,0), (0,0,1), (1,1,1)\}$ は符号である.

定義 4. $\overrightarrow{x} = (x_1, x_2, \ldots, x_n)$, $\overrightarrow{y} = (y_1, y_2, \ldots, y_n)$ に対して $x_i \neq y_i$ である i の個数を $\overrightarrow{x}$ と $\overrightarrow{y}$ の 距離 とよぶ. このときこの距離を $d(\overrightarrow{x}, \overrightarrow{y})$ と表すことにする.

例 5. $\overrightarrow{x} = (1,0,1,1,0)$, $\overrightarrow{y} = (1,0,1,0,1)$ のとき距離 $d(\overrightarrow{x}, \overrightarrow{y})$ を計算せよ.

定義 6. 符号 C に含まれる全ての異なるベクトル $\overrightarrow{x} = (x_1, x_2, \ldots, x_n)$, $\overrightarrow{y} = (y_1, y_2, \ldots, y_n)$ の距離 $d(\overrightarrow{x}, \overrightarrow{y})$ の中で最小のものを C の 最小距離 とよび, $d(C)$ で表す.

例 7. 上の例で考えた C_1 の最小距離を求めよ.

定義 8. $[n, k, d]$ 符号とは $\mathbb{F}_2^n$ の 2^k 個のベクトルをもつ符号で, 最小距離が d であるものをさす.

3　符号化, 復号法

通信路のモデル

$$\underline{\text{情報}} \underset{\text{符号化}}{\Longrightarrow} \underline{\text{送信ベクトル}} \underset{\text{通信路}}{\overset{\text{誤り}\downarrow}{\Longrightarrow}} \underline{\text{受信ベクトル}} \underset{\text{復号}}{\Longrightarrow} \underline{\text{ベクトル (情報)}}$$

- 符号化

  ある種の情報に符号 C のベクトルを対応させる. 例えば, アルファベットの a から p までを使った文字 (または文字列) を送信したい場合, アルファベットを e_7 のベクトル１６個に対応させる.

- 復号法

  A 受信ベクトルが C のベクトルであれば誤りなく受信したと考える.

  B 受信ベクトル $\overrightarrow{x}$ が C のベクトルでない場合は, $d(\overrightarrow{x}, \overrightarrow{c})$ が最小となる C のベクトル $\overrightarrow{c}$ を探して $\overrightarrow{x}$ を $\overrightarrow{c}$ に訂正する.

伝言ゲームでやったことと比べてみよう！

定理 9. C を $[n, k, d]$ 符号とする (定義 8 参照). C は $[(d-1)/2]$ 個の誤りを訂正可能. ただし, $[x]$ は x を越えない最大の整数を表す.

証明 $S(\overrightarrow{x}) = \{\overrightarrow{a} \in \mathbb{F}_2^n \mid d(\overrightarrow{x}, \overrightarrow{a}) \leq [(d-1)/2]\}$ とする. C の全ての異なるベクトル $\overrightarrow{x}, \overrightarrow{y}$ に対して $S(\overrightarrow{x})$ と $S(\overrightarrow{y})$ は交わらないことを示せばよい.

□

(注意) ここで $\overrightarrow{x}$ に対して $d(\overrightarrow{x},\overrightarrow{y}) \leq [(d-1)/2]$ となる符号のベクトル $\overrightarrow{y}$ は一つしかない (考えてみよう！).

定理 10. e_7 は $[7,4,3]$ 符号で 1 つの誤りが訂正出来る.

証明 $n=7$ は明らか. e_7 は 16 個のベクトルを含むので $k=4$. 最小距離 d が 3 であることは全ての異なる２つのベクトルの距離を計算することで確かめられる (やってみよう！). 定理 9 により $[(3-1)/2]$ 個の誤りが訂正出来る. □

n と k を固定したときには d が大きな符号が誤りの訂正能力が高くて良い符号であることが分かる. しかしながら, 次の定理から d には上限があることが分かる.

定理 11 (ハミングの限界式). C を $[n,k,d]$ 符号とすると次が成り立つ

$$\left(\sum_{i=0}^{[(d-1)/2]} {}_nC_i\right) \leq 2^{n-k}, \qquad \text{ただし } {}_nC_i = \frac{n!}{(n-i!)i!}.$$

証明 ある固定したベクトル $\overrightarrow{x}$ から $d(\overrightarrow{x},\overrightarrow{y}) \leq [(d-1)/2]$ となるベクトル $\overrightarrow{y}$ の個数は

$${}_nC_0 + {}_nC_1 + {}_nC_2 + \cdots + {}_nC_{[(d-1)/2]}$$

で与えられる. これらのベクトルは別の C のベクトルとの距離は $[(d-1)/2]$ 以下とはならないので

$$2^k({}_nC_0 + {}_nC_1 + {}_nC_2 + \cdots + {}_nC_{[(d-1)/2]}) \leq 2^n$$

を得る. □

未解決問題 12. $[72,36,16]$ 符号で全てのベクトル $\overrightarrow{x}$ に対して $d(\overrightarrow{0},\overrightarrow{x})$ が 4 の倍数になるようなものは存在するか？

最後に：
大学に入学後に 1 年で「線形代数」を習います. 今日, 講義をした内容は, 「線形代数」とその後に学ぶ群論や代数学の基礎に非常に関係しています.

4　ちょっとだけオマケ (ISBN 番号)

ほとんどの本には、裏表紙あたりに ISBN 番号が書かれています. ISBN 番号は世界中で一つの本に一つだけ割り振られている１０桁の番号なんです.
では, この ISBN 番号がどのように符号と関係しているかというと, ISBN 番号をベクトルだと思って $(x_1, x_2, x_3, \dots, x_{10})$ だったとします. そうすると

$$1 \times x_1 + 2 \times x_2 + 3 \times x_3 + \cdots + 10 \times x_{10}$$

が必ず 11 の倍数になっています, つまり $(x_1, x_2, x_3, \dots, x_{10})$ と $(1, 2, 3, \dots, 10)$ の内積が必ず 11 の倍数になっているということです.
例えば, １０月からの数理科学科の２年生向けに開講する授業「群論入門」では, 「代数学, 永尾 汎著 (朝倉書店)」を教科書に使います. この本には 4-254-11434-6 という ISBN 番号が割り振られています. 上の式に当てはめると 44 と結果が得られます.
ここで考えている符号は (詳しく定義すると準備が大変なので書けませんが) 位数 11 の有限体 $\mathbb{F}_{11}$ 上の符号で, ISBN 番号は $\mathbb{F}_{11}^{10}$ のある特別な部分集合に含まれているとみなせます. 上の計算結果が必ず 11 の倍数になっていることから, 一つだけ番号を間違えても別の本の ISBN 番号にはならない保証が得られます. 誤りを訂正することは出来ませんが誤りを検出することは可能になっています.
この他にも皆さんの身のまわりにはたくさん符号理論の基本的な概念が用いられていることがあると思います. 時間があったら是非探して見て下さい.