講義「情報理論」

第12回　通信路符号化法(1)

情報理工学専攻 情報知識ネットワーク研究室
喜田拓也

講義資料

2017/7/25

通信路符号化定理（おさらい）

定理7.4［通信路符号化定理］

通信路容量Cである通信路に対し，$R < C$ であれば，情報速度Rの符号で復号誤り率がいくらでも小さいものが存在する．しかし，$R > C$ であれば，そのような符号は存在しない．

※ Shannonの第2符号化定理とも呼ばれる

通信路容量を超えない情報速度でなら，
いくらでも精度よく通信できる
ような符号法がある！！

※でも具体的な符号の構成方法は分かっていない・・・

通信システム全体としての情報伝達の限界（おさらい）

定理7.6

情報速度$\boldsymbol{\mathcal{R}}$（ビット／秒）で発生する情報を通信路容量$\mathbf{C}$（ビット／秒）の通信路を介して送るとき，

$$\boldsymbol{\mathcal{R}} < \mathbf{C}$$

であれば，任意に小さい誤り率で情報を伝送できる．また，

$$\boldsymbol{\mathcal{R}} > \mathbf{C}$$

であれば，情報源の速度･ひずみ関数が

$$\boldsymbol{\mathcal{R}}(D_*) = \mathbf{C} \text{（ビット／秒）}$$

を満たすD_*に対し，D_*に任意に近いひずみで情報を伝送できるが，D_*より小さい平均ひずみでは伝送できない．

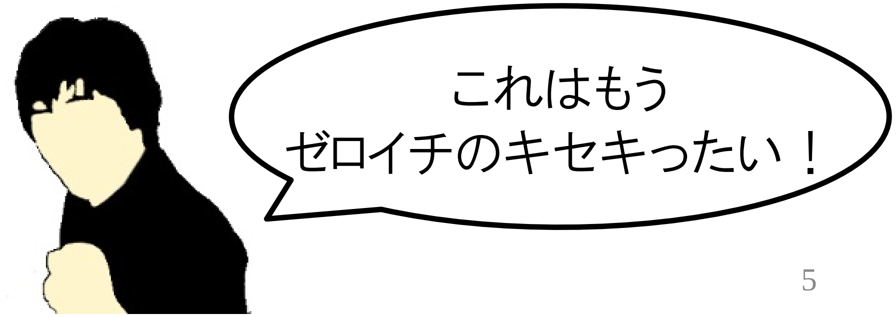
今日の内容

8.1 線形符号の基礎

8.2.1 (7,4)ハミング符号

クイズ・ザ・パリティ！

セーフ	アウト！
0	1
11	10
110	111
1000001	1001001
01010101	01101101
1111111111	11111111111



単一パリティ検査符号

長さkの系列$x_1 x_2 \cdots x_k$を記憶のない2元通信路で伝送する
そのうちの1個に誤りが生じた場合に，それを受信側で検出するにはどうすればよいだろうか？

単一パリティ検査符号の符号化方法

全体で「1」の個数が偶数になるように，末尾に記号を追加する

$$c = x_1 \oplus x_2 \oplus \cdots \oplus x_k$$

$$\boldsymbol{w} = x_1 x_2 \cdots x_k c$$

含まれる1の個数が
偶数の場合は $c = 0$
奇数の場合は $c = 1$

元の情報を表す記号$x_1, x_2, \cdots, x_k$ を情報記号という
（2元記号の場合は情報ビットと呼ばれる）
付加された記号c を検査記号（または，検査ビット）という

2番目の式はしばしば $\boldsymbol{w} = (x_1, x_2, \cdots, x_k, c)$ と表される

Try 問8.1

単一パリティ検査符号の誤り検出

長さ $k=2$ の単一パリティ検査符号は，$\boldsymbol{w}=(x_1, x_2, c)$ が

$$(0,0,0),\ (0,1,1),\ (1,0,1),\ (1,1,0)$$

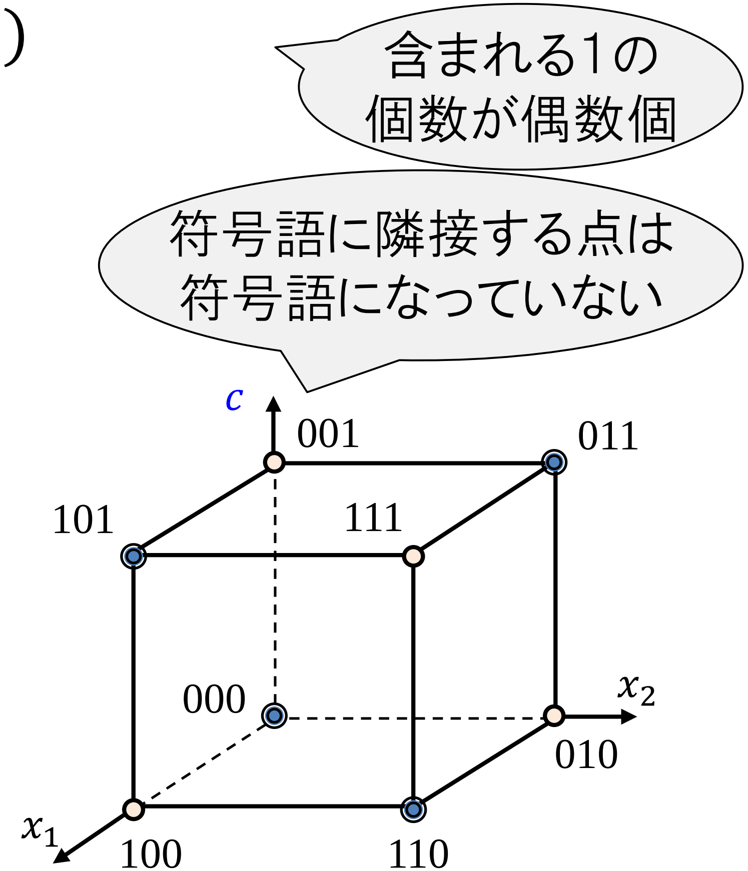
となるので，長さ3，符号語数4 の符号

$$C=\{000, 011, 101, 110\}$$

を用いているとみなせる.

一般には，長さ k の系列に対して，
長さが$k+1$で1の個数が偶数となる
2元系列すべてを符号語として用いる.
（よって符号語数は $M=2^k$個）

単一パリティ検査符号は，$\boxed{}$の誤りを検出できる！このように誤りの検出が行える符号を$\boxed{}$と呼ぶ



単一パリティ符号($k=2$)の
幾何的表現

組織符号

k個の情報記号に対して，何らかの方法で検査記号を求め，それを付加することで信頼性を高める符号長 n の等長符号を組織符号と呼ぶ

$$\boldsymbol{w} = \underbrace{x_1 x_2 \cdots x_k \, c_1 c_2 \cdots c_{n-k}}_{n}$$

符号長nで情報記号の数がkの組織符号を (n,k) 符号と書く

(n,k) 符号の効率η は，定義より

$$\eta = \frac{R}{R_{max}} = \left(\frac{\log_2 M}{n}\right) \Big/ \left(\frac{\log_2 2^n}{n}\right) = \frac{k}{n}.$$

先の例は，
(3,2)パリティ
検査符号

長さnの系列で
k個の情報記号を
送るとみれる

単一パリティ検査符号は $(k+1,k)$ 符号であり，効率は $\eta = k/(k+1)$ である

k を大きくとれば効率は上がるが，
冗長度が低くなり信頼性は小さくなる

(9,4)水平垂直パリティ検査符号

右図のように4個の情報ビットを 2×2 の配列に並べ，各行と各列に一つずつ検査ビットを付け加える.

$$c_1 = x_{11} + x_{12} \quad c_2 = x_{21} + x_{22}$$

$$c_3 = x_{11} + x_{21} \quad c_4 = x_{12} + x_{22}$$

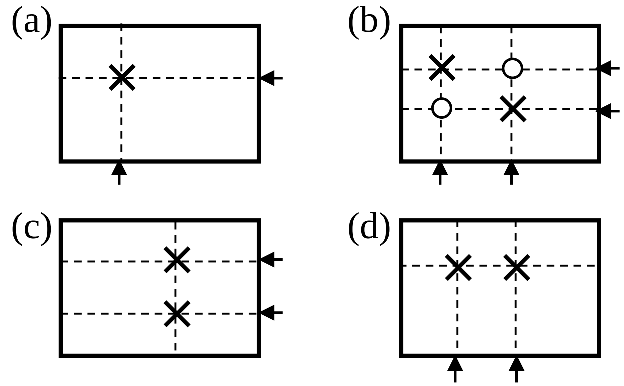
検査ビットの行の1の数が偶数になるように，検査ビットの検査ビットを右隅におく.

$$c_5 = c_1 + c_2 = x_{11} + x_{12} + x_{21} + x_{22}$$
$$= c_3 + c_4$$

この符号により $\boxed{\quad}$ 個の誤りが訂正できる.
また, $\boxed{\quad}$ 個の誤りを検出することができる.
このような符号を, $\boxed{\qquad\qquad\qquad\qquad}$
あるいは単に $\boxed{\qquad\qquad\qquad}$ と呼ぶ.

x_{11}	x_{12}	c_1
x_{21}	x_{22}	c_2
c_3	c_4	c_5

水平垂直パリティ検査符号



単一誤りの訂正と2重誤りの検出

Try 問8.3

線形符号

式 $c = x_1 + x_2 + \cdots + x_k$ のように※，検査記号が情報記号の線形な式で与えられる符号を線形符号と呼ぶ

線形符号では，任意の二つの符号語について，成分ごとの和（排他的論理和）をとると，それもまた符号語になる．この性質は線形符号となるための必要十分条件である

単一パリティ検査符号$C = \{000, 011, 101, 110\}$は線形符号
この二つの符号語 011 と 101 の和をとってみよう

$$(0, 1, 1) + (1, 0, 1) = (0 + 1, 1 + 0, 1 + 1) = (1, 1, 0)$$

※ 元は$\oplus$（排他的論理和）．この演算は，mod 2の演算，つまり2で割って余りをとる演算を考えれば簡単に+演算子で表すことができる．以降は基本的に+は$\oplus$を表す．mod 2の世界では，加算も減算も同じ意味となる．たとえば，$1 - 1 = 1 + 1 = 0$だし，$0 - 1 = 0 + 1 = 1$である．すなわち，$-$演算子もそのまま+演算子に置き換えられる．

パリティ検査方程式

(n,k)線形符号において，各検査ビット$(w_{k+1},w_{k+2},\dots,w_n)$は

$$\begin{cases} w_{k+1} = a_{11}w_1 + a_{12}w_2 + \cdots + a_{1k}w_k, \\ w_{k+2} = a_{21}w_1 + a_{22}w_2 + \cdots + a_{2k}w_k, \\ \quad\vdots \\ w_n = a_{(n-k)1}w_1 + a_{(n-k)2}w_2 + \cdots + a_{(n-k)k}w_k\ . \end{cases}$$

これを移行して得られる次の式をパリティ検査方程式と呼ぶ.

$$\begin{cases} a_{11}w_1 + a_{12}w_2 + \cdots + a_{1k}w_k + w_{k+1} = 0, \\ a_{21}w_1 + a_{22}w_2 + \cdots + a_{2k}w_k + w_{k+2} = 0, \\ \quad\vdots \\ a_{(n-k)1}w_1 + a_{(n-k)2}w_2 + \cdots + a_{(n-k)k}w_k + w_n = 0. \end{cases}$$

単一パリティ検査符号の符号語を $\boldsymbol{w} = (w_1,w_2,\dots,w_n)$ とすると,
そのパリティ検査方程式は$w_1 + w_2 + \cdots + w_{n-1} + w_n = 0$.

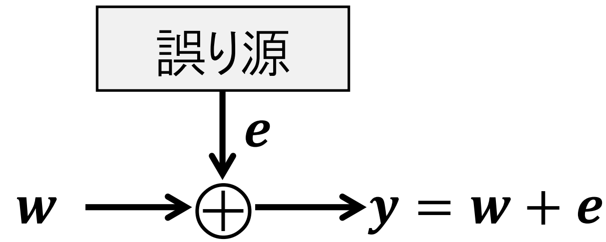
$\boldsymbol{w}$ に含まれる1の数が偶数

誤りパターンとシンドローム

ある線形符号Cの符号語$\boldsymbol{w}$ を送って$\boldsymbol{y}$が受信されたとする．これは，$\boldsymbol{w}$に誤り$\boldsymbol{e}$が加わったものと見ることができる．

$$\boldsymbol{y} = \boldsymbol{w} + \boldsymbol{e}$$

この $\boldsymbol{e}$ を誤りパターンと呼ぶ．



誤りパターン $\boldsymbol{e}$ を
用いた通信路のモデル

パリティ検査方程式の左辺に受信語$\boldsymbol{y} = (y_1, y_2, \dots, y_n)$ を(y_iをx_iとして)代入して得られる値(の組)をシンドロームと呼ぶ．

単一パリティ検査符号の場合，$\boldsymbol{w} = (w_1, w_2, \dots, w_n)$はパリティ検査方程式$x_1 + x_2 + \cdots + x_n = 0$を満たすので，

$$s = w_1 + e_1 + \cdots + w_n + e_n = e_1 + e_2 + \cdots + e_n.$$

よって，奇数個の誤りが生じた場合には$s = 1$となる．このように，シンドロームは誤りが発生した場合の有力な手がかりとなる．

ちょっと休憩

(7,4)ハミング符号(Hamming codes)

水平垂直パリティ検査符号は情報ビットよりも検査ビットが多く，あまり効率がよくない
行と列の検査による誤り位置推定という考えを一般化し，効率のよい符号を構成したい！

4個の情報ビット x_1, x_2, x_3, x_4 に対して，

$$c_1 = x_1 \qquad\quad + x_3 + x_4$$
$$c_2 = x_1 + x_2 + x_3$$
$$c_3 = \qquad +x_2 + x_3 + x_4$$

として検査ビット c_1, c_2, c_3 を作り，組織符号化（$\boldsymbol{w} = (x_1, x_2, x_3, x_4, c_1, c_2, c_3)$ ）する.
この符号は (7,4)ハミング符号と呼ばれる.
この符号は情報ビットが4であるから，符号語は $2^4 = 16$個ある.

(7,4)ハミング符号

x_1	x_2	x_3	x_4	c_1	c_2	c_3
0	0	0	0	0	0	0
1	0	0	0	1	1	0
0	1	0	0	0	1	1
1	1	0	0	1	0	1
0	0	1	0	1	1	1
1	0	1	0	0	0	1
0	1	1	0	1	0	0
1	1	1	0	0	1	0
0	0	0	1	1	0	1
1	0	0	1	0	1	1
0	1	0	1	1	1	0
1	1	0	1	0	0	0
0	0	1	1	0	1	0
1	0	1	1	1	0	0
0	1	1	1	0	0	1
1	1	1	1	1	1	1

(7,4)ハミング符号のシンドローム

符号語を$\boldsymbol{w} = (w_1, w_2, \cdots, w_7)$とすると，パリティ検査方程式は，

$$\begin{aligned} w_1 \quad\quad\quad + w_3 + w_4 + w_5 \quad\quad\quad\quad &= 0, \\ w_1 + w_2 + w_3 \quad\quad\quad\quad\quad + w_6 \quad\quad &= 0, \\ w_2 + w_3 + w_4 \quad\quad\quad\quad\quad + w_7 &= 0. \end{aligned}$$

受信語$\boldsymbol{y} = (y_1, y_2, \cdots, y_7)$に対するシンドロームは

$$\begin{aligned} s_1 &= y_1 \quad\quad\quad + y_3 + y_4 + y_5, \\ s_2 &= y_1 + y_2 + y_3 \quad\quad\quad\quad\quad + y_6, \\ s_3 &= \quad\quad\quad y_2 + y_3 + y_4 \quad\quad\quad\quad\quad + y_7. \end{aligned}$$

これは結局，誤りパターン$\boldsymbol{e} = (e_1, e_2, \cdots, e_7)$のみから次のようになる．

$$\begin{aligned} s_1 &= e_1 \quad\quad\quad + e_3 + e_4 + e_5, \\ s_2 &= e_1 + e_2 + e_3 \quad\quad\quad\quad\quad + e_6, \\ s_3 &= \quad\quad\quad e_2 + e_3 + e_4 \quad\quad\quad\quad\quad + e_7. \end{aligned}$$

同じパターンがない！

誤りパターン							シンドローム		
e_1	e_2	e_3	e_4	e_5	e_6	e_7	s_1	s_2	s_3
1	0	0	0	0	0	0	1	1	0
0	1	0	0	0	0	0	0	1	1
0	0	1	0	0	0	0	1	1	1
0	0	0	1	0	0	0	1	0	1
0	0	0	0	1	0	0	1	0	0
0	0	0	0	0	1	0	0	1	0
0	0	0	0	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0

(7,4)ハミング符号の復号（例題8.1）

先のハミング符号で符号化された系列1101111を受け取った.
この系列を復号しよう. ただし, 誤りは高々1か所だけとする.
まず, 受信語$\boldsymbol{y}=(1,1,0,1,1,1,1)$のシンドロームを計算する.

$$s_1 = y_1 + y_3 + y_4 + y_5 = 1 + 0 + 1 + 1 = 1,$$
$$s_2 = y_1 + y_2 + y_3 + y_6 = 1 + 1 + 0 + 1 = 1,$$
$$s_3 = y_2 + y_3 + y_4 + y_7 = 1 + 0 + 1 + 1 = 1.$$

これより, シンドロームは$(1,1,1)$なので,
誤りパターンは$(0,0,1,0,0,0,0)$と分かる.
3ビット目を修正して, $\boldsymbol{y}=(1,1,1,1,1,1,1)$.
したがって, 元の情報は$(1,1,1,1)$ となる.

誤りパターン							シンドローム		
e_1	e_2	e_3	e_4	e_5	e_6	e_7	s_1	s_2	s_3
1	0	0	0	0	0	0	1	1	0
0	1	0	0	0	0	0	0	1	1
0	0	1	0	0	0	0	1	1	1
0	0	0	1	0	0	0	1	0	1
0	0	0	0	1	0	0	1	0	0
0	0	0	0	0	1	0	0	1	0
0	0	0	0	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0

Try 練習問題8.1

今日のまとめ

8.1　線形符号の基礎
- 8.1.1 単一パリティ検査符号
- 8.1.2 組織符号と線形符号
- 8.1.3 水平垂直パリティ検査符号

8.2 ハミング符号
- 8.2.1 (7,4)ハミング符号

次回
- 一般のハミング符号とハミング符号の誤り訂正能力