Lưu ý: Chỉ mang tính tham khảo, vẫn còn thiếu nhiều

"Sản phẩm không phải là thuốc, không có tác dụng thay thế thuốc chữa bệnh"

1) Lý thuyết xác suất thống kê

2) Lượng tin riêng $I(x) = -\log_2 p(x)$ (bit) (chỉ dùng cơ số 2 nên không cần ghi cơ số)

3) Entropy (lượng thông tin trung bình) $H(X) = -\sum_{i=1}^{n} p(x_i).\log p(x_i)$ (bit / bản tin)

$H_{min} = 0$ khi chỉ có 1 bản tin, hay $p(x_i) = 1$, $p(x_j) = 0$ với $j \neq i$

$H_{max} = \log N$ khi $p(x_1) = p(x_2) = \cdots = p(x_N) = \frac{1}{N}$

4) Tốc độ thông tin (lượng tin trung bình trong 1s) $R = \frac{H}{T} = rH$ (bit/s)

r (tin/s), $T = \frac{1}{r}$: thời gian gửi 1 bản tin

5) Kênh rời rạc không nhớ $[P(Y)] = [P(X)].[P(Y|X)]$

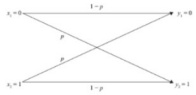
$[P(X,Y)] = [P(X)]_d[P(Y|X)]$

Nếu biểu diễn X dưới dạng ma trận đường chéo:

$$[P(X)]_d = \begin{bmatrix} P(x_1) & 0 & \dots & 0 \\ 0 & P(x_2) & \dots & 0 \\ 0 & 0 & \dots & P(x_m) \end{bmatrix}$$

Các kênh đặc biệt: Không tổn thất, xác định, không nhiễu, nhị phân đối xứng (*)

$$[P(Y|X)] = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$



6) Entropy có điều kiện

H(X): sự không chắc chắn trung bình của đầu vào kênh

H(Y): sự không chắc chắn trung bình của đầu ra kênh

H(X,Y): sự không chắc chắn trung bình của kênh truyền tổng thể

$[X|y_k] = [x_1|y_k \quad x_2|y_k \quad \dots \quad x_m|y_k]$

Khi đó: $P(X|y_k) = [P(x_1|y_k) \quad P(x_2|y_k) \dots P(x_m|y_k)]$

$$= \left[\frac{P(x_1,y_k)}{P(y_k)} \quad \frac{P(x_2,y_k)}{P(y_k)} \quad \dots \frac{P(x_m,y_k)}{P(y_k)}\right]$$

Do $P(x_1, y_k) + P(x_2, y_k) + \dots + P(x_m, y_k) = P(y_k)$

Nên: $\sum_{j=1}^{m} P(x_j|y_k) = 1$

Vậy $[X|y_k]$ được coi là một tập xác suất hoàn chỉnh.

$$H(X|Y) = -\sum_{j=1}^{m}\sum_{k=1}^{n} P(x_j, y_k).\log P(x_j|y_k)$$

Tương tự:

$$H(Y|X) = -\sum_{j=1}^{m}\sum_{k=1}^{n} P(x_j, y_k).\log P(y_k|x_j)$$

Ý nghĩa của entropy có điều kiện

$H(X|Y)$ và $H(Y|X)$ là entropy có điều kiện trung bình hay gọi đơn giản là entropy có điều kiện.

$H(X|Y)$ đo sự không chắc chắn trung bình còn lại về đầu vào kênh sau khi đã quan sát đầu ra kênh.

$H(Y|X)$ là sự không chắc chắn trung bình của đầu ra kênh khi X đã được truyền; nó cung cấp phép đo sai lỗi hoặc nhiễu.

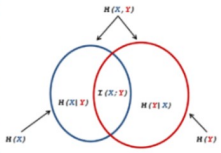
Vậy có 5 đại lượng entropy:

H(X), H(Y), H(X,Y); H(X|Y) và H(Y|X)

7) Thông tin tương hỗ I(X,Y)

phải nhớ lược đồ Venn

1. $I(X;Y) \geq 0$
2. Thông tin tương hỗ có tính chất đối xứng; $I(X;Y) = I(Y;X)$
3. $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$
4. $I(X;Y) = H(X) + H(Y) - H(X,Y)$



Lược đồ Venn

8) Dung lượng kênh

Dung lượng kênh của một kênh không nhớ rời rạc được định nghĩa là maximum của thông tin tương hỗ. Vì vậy, dung lượng kênh C được xác định bởi biểu thức:

$$C = \max_{\{P(x_j)\}} I(X;Y) \text{ bit/bản tin}$$

Nếu có r bản tin được truyền trong 1s, tốc độ truyền tin cực đại trong 1s là rC(bit/s). Đại lượng này được gọi là dung lượng kênh trong 1s hay khả năng thông qua của kênh:

$$C' = rC \text{ (bit/s)}$$

Kênh không tổn thất: $C = \max_{\{P(x_j)\}} I(X;Y) = \log_2 M$ (bit/bản tin)

Kênh xác định: $C = \max_{\{P(x_j)\}} I(X;Y) = \max_{\{P(x_j)\}} H(Y) = \log_2 n$ (bit/bản tin)

Kênh không nhiễu: $C = \max_{\{P(x_j)\}} I(X;Y) = \log_2 M = \log_2 n$ (bit/bản tin)

Kênh nhị phân đối xứng (BSC):

$$C = \max_{\{P(x_j)\}} I(X;Y) = 1 + p\log p + (1-p)\log(1-p) \text{ (bit/bản tin)}$$

9) Định lí mã hoá kênh của Shannon

(1) Cho một nguồn không nhớ rời rạc X với entropy $H(X)$ và tốc độ thông tin $r_s = \frac{1}{T_s}$ (bản tin/s). Với một kênh không nhớ rời rạc có dung lượng kênh C và tốc độ truyền $r_c = \frac{1}{T_c}$ (bản tin/s) thì nếu:

$$\frac{H(X)}{T_s} \le \frac{C}{T_c} \text{ hay } r_s . H(X) \le r_c C$$

thì luôn tồn tại một hệ thống mã hóa sao cho các bản tin được truyền qua kênh có nhiễu và có thể khôi phục lại với xác suất lỗi nhỏ tùy ý.

(2) Ngược lại nếu $\frac{H(X)}{T_s} \ge \frac{C}{T_c}$, không thể truyền tin qua kênh và khôi phục lại nó với xác suất lỗi nhỏ tùy ý.

10) Kênh liên tục

Entropy vi phân

- Xét nguồn liên tục X có hàm mật độ xác suất (pdf) là W(x).
- Nhắc lại:
  - $\int_{-\infty}^{\infty} W(x)dx = 1$
  - $m = \int_{-\infty}^{\infty} x.W(x)dx$ - Trung bình của nguồn X
  - $\sigma_X^2 = \int_{-\infty}^{\infty}(x-m)^2 W(x)dx$ – Phương sai của nguồn X
  - Khi nguồn X có trung bình m = 0 thì: $\sigma_X^2 = \int_{-\infty}^{\infty} x^2 W(x)dx$

$$h(X) = -\int_{-\infty}^{\infty} \mathrm{W}(x)\log \mathrm{W}(x)dx = \int_{-\infty}^{\infty} \mathrm{W}(x)\log\frac{1}{\mathrm{W}(x)}dx$$

Entropy vi phân có điều kiện

- X và Y là các biến ngẫu nhiên với hàm mật độ xác suất kết hợp là $f(x,y)$ và hàm mật độ xác suất biên $f(x), f(y)$.
- Entropy có điều kiện trung bình của biến ngẫu nhiên liên tục X với điều kiện Y được định nghĩa là:
- $h(X|Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log f(x|y)dxdy$ $H(X|Y) = -\sum_{j=1}^{m}\sum_{k=1}^{n} P(x_j,y_k).\log P(x_j|y_k)$
- Tương tự:
- $h(Y|X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log f(y|x)dxdy$ $H(Y|X) = -\sum_{j=1}^{m}\sum_{k=1}^{n} P(x_j,y_k).\log P(y_k|x_j)$
- Thông tin tương hỗ:
- $I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log\frac{f(x,y)}{f(x)f(y)}dxdy$

11) Gauss

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)}$

Khi đó entropy vi phân của nhiễu Gauss là:

$h(X) = -\int_{-\infty}^{\infty} f(x)\log f(x)dx = \int_{-\infty}^{\infty} f(x)\log(\frac{1}{f(x)})dx$

$= \log\sqrt{2\pi\sigma^2 e}$ bit/bản tin

KHẢ NĂNG THÔNG QUA CỦA KÊNH GAUSS

- *Khả năng thông qua của kênh Gauss* trong trường hợp tín hiệu vào là hàm liên tục của thời gian liên tục với phổ hữu hạn F là:

$$C' = F\log\left(1+\frac{P}{N_0 F}\right) = F\log\left(1+\frac{P}{P_n}\right)$$

Với N_0 : mật độ phổ công suất thực tế của nhiễu

P_n: công suất trung bình của nhiễu trong dải tần F.

P: công suất trung bình của tín hiệu hữu ích (tín hiệu phát)

- Nhận xét:
  - Khi $F \to 0$ thì $C' \to 0$
  - Khi $F \uparrow$ thì $C' \uparrow$
  - Khi $F \to \infty$ thì $C'_\infty = 1{,}443.\frac{P}{N_0}$

12) Mã hoá nguồn $H(X) \le L < H(X) + \frac{1}{n}$

Độ dài trung bình từ mã: Xét một nguồn rời rạc không nhớ X có entropy hữu hạn $H(X)$ và tập tin $\{x_1, x_2, \dots, x_m\}$ với xác suất xảy ra tương ứng là $P(x_j)$ .Nếu từ mã nhị phân được ấn định cho ký tự x_j có độ dài n_j bit, độ dài trung bình từ mã L cho mỗi ký tự nguồn được định nghĩa là:

$$L = \sum_{j=1}^{m} P(x_j)\, n_j \text{ (bit/ký tự)}$$

Hiệu suất mã: Hiệu suất mã được cho bởi: $\eta = \frac{L_{min}}{L}$

Ở đó L_{min} là giá trị nhỏ nhất có thể có của L.

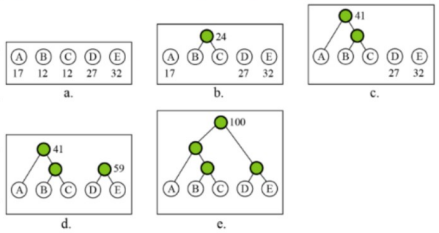
Độ dư thừa mã: Độ dư thừa γ của mã được định nghĩa là: $\gamma = 1 - \eta$

12) Mã Huffman

Chọn 2 cái có số tần xuất hiện ít nhất rồi ghép làm 1, lặp đến hết (giống câu nói dây codeptit)

```
priority_queue <int, vector<int>, greater<int>> q;
while (q.size() > 1) {
    int a = q.top(); q.pop();
    int b = q.top(); q.pop();
    q.push(a + b);
}
```

Ký tự	A	B	C	D	E
Tần suất xuất hiện	17	12	12	27	32

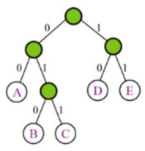


Cho tập 5 ký tự A, B, C, D, E với tần suất xuất hiện tương ứng trong bảng dưới đây.

A. Thực hiện mã hoá Huffman cho tập ký tự này.

B. Tính hiệu suất mã

Ký tự	A	B	C	D	E
Tần suất xuất hiện	17	12	12	27	32
	0,17	0,12	0,12	0,27	0,32



A: 00 D: 10
B: 010 E: 11
C: 011
Code

Ký tự	A	B	C	D	E
Tần suất xuất hiện	17	12	12	27	32

(quy ước nhánh bên nào là 0 hay 1 đều được, miễn cùng bên thì phải giống nhau)

Eg: Tất cả nhánh trái cùng là 0, phải là 1 như trên VD

13) Mã Khối

Bản tin gồm k bit thông tin → có 2^k từ mã

Định nghĩa 5.1. Một từ mã là một chuỗi các ký tự.

Định nghĩa 5.2. Một bộ mã là một tập các vector gọi là từ mã

Định nghĩa 5.3. Trọng số của một từ mã bằng số lượng các thành phần khác 0 trong từ mã. Trọng số của từ mã c được ký hiệu là $w(c)$.

Định nghĩa 5.4. Khoảng cách Hamming giữa hai từ mã là số các vị trí mà ở đó hai từ mã khác nhau. Ký hiệu khoảng cách Hamming giữa hai từ mã c_1 và c_2 là $d(c_1, c_2)$.

Tính chất: $d(c_1, c_2) = w(c_1 + c_2)$

Khoảng cách tối thiểu của bộ mã: $d_0 = \min d(c_i, c_j)$

Định nghĩa 5.5. Một bộ mã khối gồm một tập các từ mã độ dài cố định. Độ dài cố định của các từ mã này được gọi độ dài khối và thường được ký hiệu là n. Vì vậy, một bộ mã độ dài n gồm một tập các từ mã có n thành phần.

Mã khối tuyến tính

Mã hệ thống tuyến tính (n, k): là mã tuyến tính độ dài từ mã n trong đó có k ký tự đầu tiên (hoặc cuối cùng) của từ mã chính là k ký tự thông tin. $(n - k)$ ký tự còn lại gọi là các ký tự kiểm tra chẵn lẻ (dư thừa).

Phần kiểm tra (dư thừa) (n-k) bit	Phần bản tin (k) bit

Tổng của hai từ mã trong bộ mã cũng là một từ mã thuộc bộ mã.
Từ mã toàn 0 luôn luôn là một từ mã.
Khoảng cách Hamming tối thiểu giữa hai từ mã của một bộ mã khối tuyến tính bằng trọng số tối thiểu của các từ mã khác 0 trong bộ mã. $d_0 = \min\{w(c_i)\}$

Định lý về khả năng phát hiện sai:
Một bộ mã khối tuyến tính (n, k, d_0) có khả năng phát hiện được t sai thỏa mãn: $t \le d_0 - 1$.

Định lý về khả năng sửa sai:
Một bộ mã khối tuyến tính (n, k, d_0) có khả năng sửa được t sai thỏa mãn: $t \le \left[\frac{d_0 - 1}{2}\right]$

- Ma trận sinh G, dạng hệ thống $G = [I|P]$ hoặc $[P|I]$
- Ma trận kiểm tra H: $G.H^T = 0$
- c: là một từ mã → $c = m.G$
  → $m.G.H^T = c.H^T = 0$

Nếu $\boldsymbol{G}$ ở dạng hệ thống, ma trận $\boldsymbol{H}$ sẽ có dạng:

$\boldsymbol{G} = (\boldsymbol{I}\|\boldsymbol{P})$	$\mathbf{H} = (\boldsymbol{P}^T\|\mathbf{I}')$
$\boldsymbol{G} = (\mathbf{P}\|\mathbf{I})$	$\mathbf{H} = (\mathbf{I}'\|\boldsymbol{P}^T)$

14) Mã cyclic (n, k, d_0)

k: số bit thông tin
d_0: khoảng cách Hamming

- Đa thức sinh $g(x)$ có $\deg(g(x)) = r = n - k$

⇒ Ma trận sinh $G = \begin{bmatrix} x^0 g(x) \\ \dots \\ x^{k-1} g(x) \end{bmatrix}$

- Đa thức kiểm tra $h(x) = \dfrac{x^n + 1}{g(x)}$, $\deg(h(x)) = k$, $h_0 = h_k = 1$

→ $h^*(x) = x^{\deg h(x)} . h(x^{-1})$

⇒ Ma trận kiểm tra $H = \begin{bmatrix} x^0 . h^*(x) \\ \dots \\ x^{r-1} . h^*(x) \end{bmatrix}$

Một số biến đổi

- $(x^n + 1)^2 = x^{2n} + 1$
- $x^n + 1 = (x + 1)(x^{n-1} + x^{n-2} + \cdots + x + 1)$

15) Phương pháp nhân và chia

recommend xem youtube, đọc slide nhé :>

Xây dựng ma trận ở dạng hệ thống

$C(n,k) \Rightarrow l = 1,2,\dots k \Rightarrow x^{n-l} = x^{n-1}, x^{n-2}, \dots, x^{n-k}$

$r_1 = x^{n-k} \bmod g(x)$

$r_2 = x^{n-k+1} \bmod g(x)$

$\dots$

$r_k = x^{n-1} \bmod g(x)$

$\Rightarrow G = \begin{bmatrix} x^{n-k} + r_1 \\ x^{n-k+1} + r_2 \\ \dots \\ x^{n-1} + r_k \end{bmatrix} = (P \mid I)$ hay $(I \mid P) \Rightarrow H = (I' \mid P^T)$ hay $(P^T \mid I)$

Đa thức bất khả quy: đa thức chỉ chia hết cho 1 và chính nó, $g_0 = g_k = 1$

Một số đa thức BKQ thường gặp:

Bậc	Đa thức
1	$1 + x$
2	$1 + x + x^2$
3	$1 + x + x^3$, $1 + x^2 + x^3$
4	$1 + x + x^4$, $1 + x^3 + x^4$, $1 + x + x^2 + x^3 + x^4$

Phân tích $x^n + 1$ thành tích các đa thức bất khả quy

$n = 2^k - 1 \rightarrow$ phân tích thành tích các đa thức BKQ có bậc là ước của k

Đa thức sinh $g(x)$ của $C(n,k)$

- $(x^n + 1) \vdots g(x)$
- $\deg(g(x)) = n - k$
- $g(x)$ là đa thức monic

Đa thức $d(x)$ là một từ mã của bộ mã

- $d(x) \vdots g(x)$
- $\deg(d(x)) < n$
- $\frac{d(x)}{g(x)}$ chứa k bit thông tin

Khoảng cách Hamming: dựa vào ma trận kiểm tra H

số cột ít nhất của H có tổng = 0 (phụ thuộc tuyến tính) là khoảng cách Hamming (d₀)

* log ở trong môn này chỉ dùng cơ số 2 nên $\log p(x)$ tức là $\log_2 p(x)$

chỉ có duy nhất đổi dB là cơ số 10: VD $SNR = 30dB = 10^{30/10} = 10^3$