C1: Tổng quan và Các khái niêm cơ bản

Lý thuyết thông tin

Biên soạn: Phạm Văn Sự

Bộ môn Xử lý tín hiệu và Truyền thông
Khoa Kỹ thuật Điện tử I
Học viện Công nghệ Bưu chính Viễn thông

ver. 22a

Notes

Mục tiêu của bài học

- Giới thiệu sơ lược về vai trò, lịch sử và hướng phát triển của môn khoa học Lý thuyết thông tin
- Cung cấp các định nghĩa, khái niệm cơ bản nền tảng của Lý thuyết thông tin
- Giới thiệu mô hình tổng quan của hệ thống truyền tin
- Những khía cạnh đánh giá một hệ thống truyền tin

Notes

Các câu hỏi cần trả lời

- Môn khoa học Lý thuyết thông tin đóng vai trò gì với các môn khoa học khác?
- Những mốc thời gian cơ bản và những thành tựu của môn khoa học Lý thuyết thông tin cho đến nay?
- Hiện nay môn khoa học Lý thuyết thông tin được phát triển thế nào? theo những hướng nào? ví dụ minh họa các hướng này?
- Ba định nghĩa nền tảng của Lý thuyết thông tin là gì? T/c của chúng?
- Sơ đồ một hệ thống truyền tin cơ bản? Vai trò, chức năng và nguyên lý của các khối trong sơ đồ?
- Để đánh giá một hệ thống truyền tin, chúng ta đánh giá theo các yếu tố nào?

Notes

C1: Tổng quan và Các khái niệm cơ bản

Nội dung chính

Notes

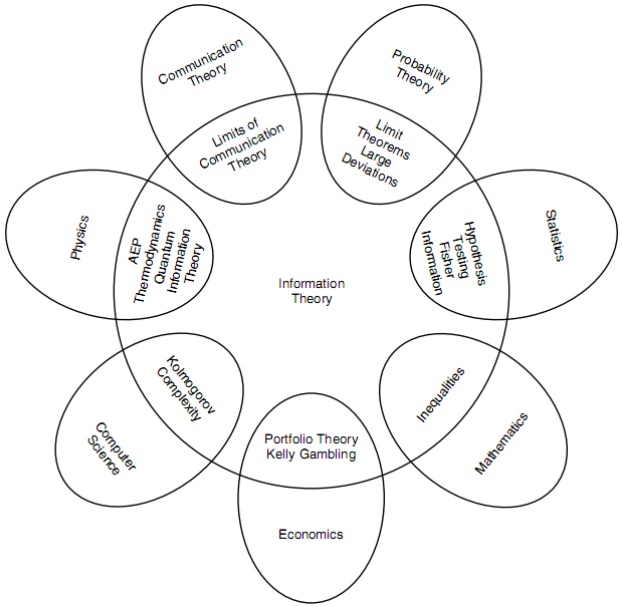
C1: Tổng quan và Các khái niêm cơ bản

Nội dung chính

Notes

Sơ lược về môn Lý thuyết thông tin

Vị trí và vai trò của Lý thuyết thông tin



Hình: © Thomas M. Cover et al

Notes

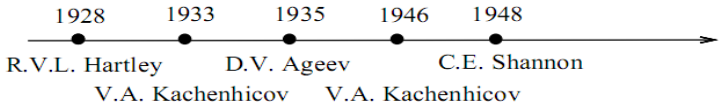
C1: Tổng quan và Các khái niệm cơ bản

Nội dung chính

Notes

Sơ lược về môn Lý thuyết thông tin

Lịch sử phát triển của môn LTTT



Notes

Notes

- Nén dữ liệu (Data compression)
  - Giới hạn dưới của độ dài trung bình của các biểu diễn thông tin?
  - Với giới hạn "méo" cho trước, tốc độ mã hóa tối đa là bao nhiêu?
- Truyền dữ liệu (Data transmission)
  - Phương thức mã hóa kênh nào để phía thu có thể thu và giải mã với xác suất lỗi nhỏ nhất?
  - Một bộ mã hóa kênh tối ưu có thể đạt được cặp giá trị (R, p_e) ở đâu?
- Thông tin mạng (Network information theory)
  - Bài toán "Nén" và "Truyền" trong mạng gồm nhiều nguồn/người dùng?
- Suy luận (Inference)
  - Điều gì/kết cục gì sẽ xảy ra tiếp?
- Đánh bạc và đầu tư (Gambling and investment)
  - Giới hạn trên của "tốc độ nhân đôi" - cực đại tiệm cận lũy thừa của sự giàu có (tăng trưởng)?
- Tính toán độ phức tạp (Complexity theory)

Notes

C1: Tổng quan và Các khái niêm cơ bản

Nội dung chính

1. Sơ lược về môn Lý thuyết thông tin (LTTT)
   - Vị trí và vai trò của Lý thuyết thông tin
   - Lịch sử phát triển của môn LTTT
   - Các bài toán thực tế của LTTT
2. Sơ đồ tổng quát của hệ thống truyền tin và các khái niệm cơ bản
   - Các khái niệm cơ bản
   - Sơ đồ tổng quát của hệ thống truyền tin
   - Sơ lược về các phương pháp xử lý thông tin trong hệ thống thông tin
   - Những tiêu chí đánh giá chất lượng một hệ thống thông tin

Notes

C1: Tổng quan và Các khái niêm cơ bản

Nội dung chính

1. Sơ lược về môn Lý thuyết thông tin (LTTT)
   - Vị trí và vai trò của Lý thuyết thông tin
   - Lịch sử phát triển của môn LTTT
   - Các bài toán thực tế của LTTT
2. Sơ đồ tổng quát của hệ thống truyền tin và các khái niệm cơ bản
   - Các khái niệm cơ bản
   - Sơ đồ tổng quát của hệ thống truyền tin
   - Sơ lược về các phương pháp xử lý thông tin trong hệ thống thông tin
   - Những tiêu chí đánh giá chất lượng một hệ thống thông tin

Notes

Sơ đồ tổng quát của hệ thống truyền tin và các khái niệm cơ bản

Các định nghĩa cơ bản (1)

Định nghĩa (Thông tin - Information)

Thông tin là những tính chất xác định của vật chất mà con người trực tiếp hoặc gián tiếp thông qua hệ thống kỹ thuật thu nhận được từ thế giới vật chất bên ngoài hoặc từ những quá trình xảy ra trong bản thân nó, nhằm mang lại sự hiểu hiểu biết về chúng.

- Khách quan
- Đa dạng

Định nghĩa (Tin - Message)

Tin là dạng vật chất cụ thể để biểu diễn hoặc thể hiện thông tin

- Tin liên tục
- Tin rời rạc

Notes

Sơ đồ tổng quát của hệ thống truyền tin và các khái niệm cơ bản

Các định nghĩa cơ bản (2)

Định nghĩa (Tín hiệu - Signal)

Tín hiệu là các đại lượng vật lý biến thiên, phản ánh tin cần truyền.

- Sự biến đổi tham số riêng của quá trình vật lý mới là tín hiệu

Notes

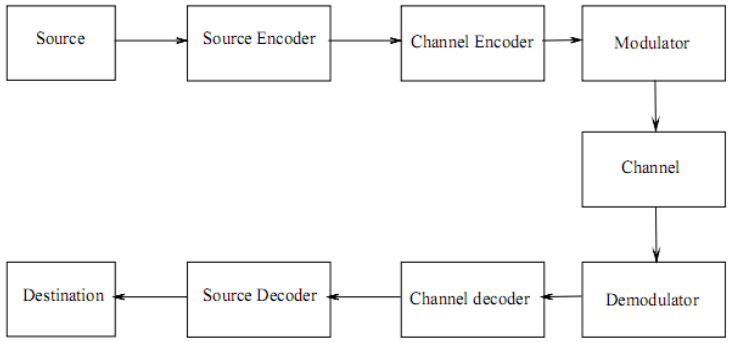
C1: Tổng quan và Các khái niêm cơ bản

Nội dung chính

Notes

Sơ đồ tổng quát của hệ thống truyền tin và các khái niệm cơ bản

Sơ đồ tổng quát của hệ thống truyền tin



Hình: Sơ đồ tổng quát hệ thống thông tin

Notes

Sơ đồ tổng quát của hệ thống truyền tin

Các khối cơ bản (1)

Định nghĩa (Nguồn tin - Source)

Nguồn là nơi sản sinh ra tin.

- Đặc tính: Nguồn liên tục, nguồn rời rạc
- Tính chất: Thống kê, hàm ý

Định nghĩa (Máy phát - Transmitter)

Máy phát là thiết bị biến đổi tập tin thành tập tín hiệu tương ứng để truyền đi.

- Phép biến đổi phải đảm bảo tính song ánh (đơn trị 2 chiều)
- Tổng quát gồm: Mã hóa và điều chế

Notes

Sơ đồ tổng quát của hệ thống truyền tin

Các khối cơ bản (2)

Định nghĩa (Kênh truyền tin - Channel)

Kênh truyền tin là tập hợp các môi trường vật lý trong đó tín hiệu được truyền đi từ nguồn đến nơi nhận tin.

- Kênh (channel) thường được hiểu là phần đường truyền tin từ phía phát đến phía thu.

Định nghĩa (Máy thu - Receiver)

Máy thu là thiết bị thu nhận tín hiệu và từ đó thiết lập lại thông tin.

- Máy thu thực hiện các phép biến đổi ngược máy phát.
- Tổng quát gồm: Giải điều chế và giải mã.

Notes

Sơ đồ tổng quát của hệ thống truyền tin

Các khối cơ bản (3)

Định nghĩa (Nhận tin - Reception)

Là việc thu nhận thông tin nhằm sao lưu, biểu thị và xử lý tin.

Định nghĩa (Nhiễu - Noise)

Là các yếu tố có ảnh hưởng xấu đến việc thu nhận tin.

Notes

C1: Tổng quan và Các khái niệm cơ bản

Nội dung chính

Notes

Sơ đồ tổng quát của hệ thống truyền tin

Sơ lược về các phương pháp xử lý thông tin trong hệ thống thông tin

- Định dạng và mã hóa nguồn - Sử dụng tối thiểu tài nguyên để biểu diễn tin một cách đầy đủ nhất: Lấy mẫu, lượng tử hóa, điều chế xung mã (PCM), PCM vi phân, mã Huffman, etc.
- Mã hóa kênh - Sử dụng tối thiểu tài nguyên để đảm bảo việc truyền nhận thông tin với lỗi ít nhất: mã hóa khối, mã hóa liên tục, etc.
- Điều chế - Truyền thông tin với tốc độ cao nhất, tốn ít năng lượng nhất: Điều chế dịch khóa pha (PSK), Điều chế dịch khóa tần (FSK), etc.
- Ghép kênh/đa truy nhập - Chia sẻ tài nguyên tốt nhất cho người dùng trong hệ thống: TDM/TDMA, CDMA, MC-CDMA, etc.
- Bảo mật - Đảm bảo tính bí mật, xác thực và toàn vẹn của tin trong quá trình truyền.

Notes

C1: Tổng quan và Các khái niệm cơ bản

Nội dung chính

Notes

Sơ đồ tổng quát của hệ thống truyền tin

Những tiêu chí đánh giá chất lượng một hệ thống thông tin

- Tính hiệu quả
  - Tốc độ truyền tin cao.
  - Truyền đồng thời nhiều tin khác nhau.
  - Chi phí cho một bít thông tin thấp.
- Độ tin cậy cao.
  - Đảm bảo độ chính xác của việc nhận thông tin.
- An toàn.
  - Bí mật (Confidentiality)
  - Xác thực (Authentication)
  - Toàn vẹn (Integrity)
  - Khả dụng (Availability)
- Đảm bảo chất lượng dịch vụ (Quality of Service - QoS).

Notes

Notes

C2: Lý thuyết thông tin thống kê

Lý thuyết thông tin

Biên soạn: Phạm Văn Sự

Bộ môn Xử lý tín hiệu và Truyền thông
Khoa Kỹ thuật Điện tử I
Học viện Công nghệ Bưu chính Viễn thông

ver. 22a

Notes

Mục tiêu của bài học

- Công thức tính và đơn vị đo lường của thông tin
- Đánh giá lượng tin trung bình thống kê của nguồn
- Mối liên hệ về lượng tin giữa các nguồn thông tin
- Lượng thông tin trung bình truyền qua kênh

Notes

Các câu hỏi cần trả lời

- Một sự kiện xuất hiện sẽ mang lại một lượng tin bằng bao nhiêu? Đơn vị lượng tin?
- Lượng tin tiên nghiệm, hậu nghiệm, tương hỗ là gì? Ý nghĩa các đại lượng trong mô hình phát - thu? Giá trị và ý nghĩa của các đại lượng này trong hai trường hợp cực đoan của kênh?
- Lượng thông tin trung bình thống kê của nguồn rời rạc không nhớ xác định thế nào? Tính chất? Áp dụng?
- Mối quan hệ về lượng tin giữa các nguồn thông tin? Tính chất? Mối quan hệ giữa các đại lượng?
- Lượng tin trung bình truyền qua kênh xác định thế nào?
- Suy diễn các khái niệm tương tự cho nguồn liên tục?

Notes

Phần I

Lý thuyết thông tin thống kê cho nguồn rời rạc

Notes

C2: Lý thuyết thông tin thống kê

Nội dung chính

1. Đo lường thông tin
   - Lượng tin riêng
   - Lượng tin hậu nghiệm, lượng tin tương hỗ, hai trạng thái cực đoan của kênh
2. Entropy và các đại lượng liên quan của nguồn rời rạc
   - Entropy
   - Entropy của các trường sự kiện đồng thời
   - Entropy có điều kiện
   - Entropy tương đối và Lượng thông tin tương hỗ giữa các nguồn
   - Tính chất và các mối quan hệ giữa các đại lượng
3. Lý thuyết thông tin thống kê cho nguồn liên tục
   - Tín hiệu liên tục, Nguồn liên tục
   - Entropy vi phân
   - Entropy vi phân hợp, Entropy vi phân có điều kiện, Lượng tin tương hỗ của các nguồn liên tục

Notes

C2: Lý thuyết thông tin thống kê

Nội dung chính

1. Đo lường thông tin
   - Lượng tin riêng
   - Lượng tin hậu nghiệm, lượng tin tương hỗ, hai trạng thái cực đoan của kênh
2. Entropy và các đại lượng liên quan của nguồn rời rạc
   - Entropy
   - Entropy của các trường sự kiện đồng thời
   - Entropy có điều kiện
   - Entropy tương đối và Lượng thông tin tương hỗ giữa các nguồn
   - Tính chất và các mối quan hệ giữa các đại lượng
3. Lý thuyết thông tin thống kê cho nguồn liên tục
   - Tín hiệu liên tục, Nguồn liên tục
   - Entropy vi phân
   - Entropy vi phân hợp, Entropy vi phân có điều kiện, Lượng tin tương hỗ của các nguồn liên tục

Notes

Đo lường thông tin

Lượng tin riêng: Ví dụ 1

Ví dụ

Chúng ta nhận được một bức thư.

- **TH1**: Đã biết hoặc đoán biết chắc chắn nội dung của bức thư → không có độ bất định: bức thư không mang lại thông tin.
- **TH2**: Không biết và có thể đoán biết không chắc chắn nội dung của bức thư → có độ bất định: bức thư mang lại một lượng thông tin.
- **TH3**: Không biết và không thể đoán biết không được nội dung của bức thư → độ bất định rất lớn: bức thư mang lại một lượng thông tin lớn.

⇒ Độ bất định: một đặc trưng quan trọng trong đo lường lượng thông tin.

- Lượng thông tin tỷ lệ thuận với độ bất định

⇒ Không có độ bất định: không có thông tin. ⇒ Lượng thông tin thu được bằng cách làm giảm độ bất định.

Notes

Đo lường thông tin

Lượng tin riêng: Ví dụ 2

Ví dụ

Một rổ đựng n bóng ($n = 1, 2, \dots$), các bóng được đánh nhãn từ 1 đến hết. Lấy ngẫu nhiên một bóng và quan sát nhãn của nó. Quan sát xác suất, độ bất định của sự kiện chúng ta lấy được một bóng có nhãn là "1".

n	Xác suất	Độ bất định
1	1	0
2	1/2	$\neq 0$
$\vdots$	$\vdots$	$\vdots$
∞	≈ 0	∞

Lượng thông tin: là một hàm giảm của xác suất xuất hiện của tin.

Notes

Đo lường thông tin

Lượng thông tin riêng: Định nghĩa

Nhận xét: Gọi x là một tin với xác suất xuất hiện $p(x)$, gọi $I(x)$ là đại lượng biểu diễn *lượng thông tin mà chúng ta thu được khi biết rằng x đã xảy ra (hoặc một cách tương đương, lượng độ bất định mất đi khi chúng ta biết x đã xảy ra)*

1. $I(x)$: là một hàm của $p(x) \Rightarrow I(x) = I(p(x))$
   - $I(.)$ là liên tục của $p(x)$ với $p(x) \in [0, 1]$; $I(p(x) = 1) = 0$.
   - $I(.)$ là một hàm đơn điệu giảm theo $p(x)$.
   - $I(x) \geq 0$.
2. Nếu x và y là hai tin độc lập thì $I(x \cap y) = I(x) + I(y)$
   - $I(p(x) \times p(y)) = I(p(x)) + I(p(y))$.

Định nghĩa: Lượng thông tin riêng

Một tin (sự kiện) x với xác suất xuất hiện $p(x)$ thì việc nó xuất hiện sẽ mang lại lượng thông tin, hay còn gọi là lượng tin riêng/lượng thông tin tiên nghiệm, được xác định bởi:

$$I(x) \triangleq -log(p(x))$$

Notes

Đo lường thông tin

Lượng thông tin riêng: Đơn vị

$$I(x_k) = -\log(p(x_k))$$

- Logarithm:
  - Cơ số 2: đơn vị [*bit*].
  - Cơ số $e = 2,7\ldots$: đơn vị [*nat*].
  - Cơ số 10: đơn vị [*hartley*].

Ví dụ

Một bình đựng 2 viên bi màu đen và ba viên bi màu trắng. Thực hiện việc lấy ngẫu nhiên hai lần liên tiếp, mỗi lần một viên bi, bi đã được lấy không được bỏ lại bình. Gọi x là thông điệp cho chúng ta biết đã lấy được cả hai viên bi màu đen. Tính lượng tin của thông điệp x.

Notes

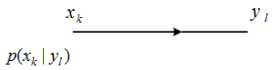
C2: Lý thuyết thông tin thống kê

Nội dung chính

Notes

Đo lường thông tin

Lượng thông tin hậu nghiệm, lượng tin tương hỗ, 2 trạng thái cực đoan của kênh



Biết đã nhận được tin $y_l \rightarrow$ tin x_k phát đi với xác suất $p(x_k|y_l)$

- $I(x_k|y_l) \triangleq -log(p(x_k|y_l))$: Lượng thông tin hậu nghiệm
  - Lượng tin riêng về x_k sau khi đã có (biết) y_l
- $I(x_k;y_l) \triangleq I(x_k) - I(x_k|y_l)$: Lượng thông tin chéo về x_k do y_l mang.
  - Lượng tin tương hỗ giữa tin x_k và y_l
- $\Rightarrow I(x_k|y_l) = I(x_k) - I(x_k;y_l)$: Lượng thông tin tổn hao trên kênh.

Nhận xét:

- Kênh không có nhiễu: $I(x_k|y_l) = 0$; $I(x_k;y_l) = I(x_k)$
- Kênh bị đứt (bị nhiễu tuyệt đối): $I(x_k;y_l) = 0$, $I(x_k|y_l) = I(x_k)$.

Notes

C2: Lý thuyết thông tin thống kê

Nội dung chính

1. Đo lường thông tin
   - Lượng tin riêng
   - Lượng tin hậu nghiệm, lượng tin tương hỗ, hai trạng thái cực đoan của kênh
2. Entropy và các đại lượng liên quan của nguồn rời rạc
   - Entropy
   - Entropy của các trường sự kiện đồng thời
   - Entropy có điều kiện
   - Entropy tương đối và Lượng thông tin tương hỗ giữa các nguồn
   - Tính chất và các mối quan hệ giữa các đại lượng
3. Lý thuyết thông tin thống kê cho nguồn liên tục
   - Tín hiệu liên tục, Nguồn liên tục
   - Entropy vi phân
   - Entropy vi phân hợp, Entropy vi phân có điều kiện, Lượng tin tương hỗ của các nguồn liên tục

Notes

C2: Lý thuyết thông tin thống kê

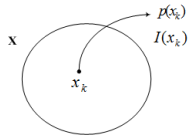
Nội dung chính

1. Đo lường thông tin
   - Lượng tin riêng
   - Lượng tin hậu nghiệm, lượng tin tương hỗ, hai trạng thái cực đoan của kênh
2. Entropy và các đại lượng liên quan của nguồn rời rạc
   - Entropy
   - Entropy của các trường sự kiện đồng thời
   - Entropy có điều kiện
   - Entropy tương đối và Lượng thông tin tương hỗ giữa các nguồn
   - Tính chất và các mối quan hệ giữa các đại lượng
3. Lý thuyết thông tin thống kê cho nguồn liên tục
   - Tín hiệu liên tục, Nguồn liên tục
   - Entropy vi phân
   - Entropy vi phân hợp, Entropy vi phân có điều kiện, Lượng tin tương hỗ của các nguồn liên tục

Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Entropy - Lượng tin trung bình thống kê của nguồn



X: nguồn rời rạc không nhớ (DMS) gồm các tin x_k xung khắc với xác suất xuất hiện $p(x_k)$

Định nghĩa (Entropy)

Entropy của nguồn rời rạc không nhớ X là trung bình thống kê của lượng thông tin riêng của các tin (phần tử) x_k (xung khắc) thuộc nguồn, ký hiệu là $H(X)$.

$$H(X) \triangleq E[I(x_k)] = \sum_{k=1}^{N} p(x_k) I(x_k) = -\sum_{k=1}^{N} p(x_k) \log(p(x_k))$$
$$= E[-\log(p(x_k))]$$

- $H(X)$ còn được gọi là entropy một chiều của nguồn rời rạc.
- $H(X)$ có đơn vị của lượng thông tin (bit, nat, hartley).

Notes

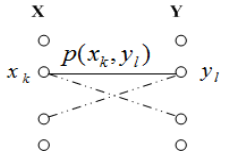
C2: Lý thuyết thông tin thống kê

Nội dung chính

Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Entropy của các trường sự kiện đồng thời - Entropy hợp



Hình: Mô hình của cặp nguồn rời rạc X và Y

Định nghĩa (Entropy hợp)

Entropy hợp $H(X,Y)$ của một cặp nguồn rời rạc (X,Y) (còn gọi là Entropy của trường sự kiện đồng thời (X,Y)) với xác suất phân bố đồng thời của các tin x_k và y_l là $p(x_k, y_l)$ được cho bởi công thức:

$$\begin{aligned} H(X,Y) &\triangleq -\sum_{x_k \in X}\sum_{y_l \in Y} p(x_k,y_l)\log(p(x_k,y_l)) = -\sum_{k=1}^{N}\sum_{l=1}^{M} p(x_k,y_l)\log(p(x_k,y_l)) \\ &= E[-\log(p(x_k,y_l))]_{(x_k,y_k)\in(X,Y)} \end{aligned}$$

Notes

C2: Lý thuyết thông tin thống kê

Nội dung chính

Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Entropy có điều kiện: Entropy có điều kiện từng phần (1/2)

Định nghĩa (Entropy có điều kiện từng phần - Partial Conditional Entropy)

Cho hai nguồn rời rạc X, Y. $H(X|Y=y_l)$ được gọi là Entropy có điều kiện từng phần, là Entropy có điều kiện về một nguồn tin này khi đã nhận được một tin nhất định của nguồn kia.

$$\begin{aligned} H(X|Y=y_l) &\triangleq E[I(x_k|Y=y_l)]_{x_k\in X|Y=y_l} \\ &= \sum_{x_k\in X} p(x_k|Y=y_l)I(x_k|Y=y_l) \\ &= -\sum_{x_k\in X} p(x_k|Y=y_l)\log(p(x_k|Y=y_l)) \\ &= -\sum_{k=1}^{N} p(x_k|y_l)\log(p(x_k|y_l)) \end{aligned}$$

$H(X|Y=y_l)$: lượng tin tổn hao trung bình của mỗi tin ở đầu phát khi đầu thu đã

Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Entropy có điều kiện: Entropy có điều kiện từng phần (2/2)

Định nghĩa (Entropy có điều kiện từng phần - Partial Conditional Entropy)

Cho hai nguồn rời rạc X, Y. $H(Y|X=x_k)$ được gọi là Entropy có điều kiện từng phần, là Entropy có điều kiện về một nguồn tin này khi đã phát đi một tin nhất định của nguồn kia.

$$\begin{aligned} H(Y|X=x_k) &\triangleq E[I(y_l|X=x_k)]_{y_l\in Y|X=x_k} \\ &= \sum_{y_l\in Y} p(y_l|X=x_k)I(y_l|X=x_k) \\ &= -\sum_{y_l\in Y} p(y_l|X=x_k)\log(p(y_l|X=x_k)) \\ &= -\sum_{l=1}^{M} p(y_l|x_k)\log(p(y_l|x_k)) \end{aligned}$$

$H(Y|X=x_k)$: lượng tin riêng trung bình chứa trong mỗi tin ở đầu thu khi đầu

Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Entropy có điều kiện (1/2)

Định nghĩa (Entropy có điều kiện)

Với một cặp nguồn rời rạc (X, Y) có xác suất phân bố hợp $p(x_k, y_l)$, xác xuất phân bố có điều kiện $p(x_k|y_l)$, Entropy có điều kiện $H(X|Y)$ được cho bởi công thức:

$$\begin{aligned} H(X|Y) &\triangleq E[H(X|Y=y_l)]_{y_l \in Y} = \sum_{y_l \in Y} p(y_l) H(X|Y=y_l) \\ &= -\sum_{y_l \in Y} p(y_l) \sum_{x_k \in X} p(x_k|y_l) \log(p(x_k|y_l)) \\ &= -\sum_{k=1}^{N} \sum_{l=1}^{M} p(x_k, y_l) \log(p(x_k|y_l)) \\ &= E[-\log(p(X|Y))]_{p(x_k,y_l)} \end{aligned}$$

$H(X|Y)$: lượng tin riêng tổn hao trung bình của mỗi tin ở đầu phát khi đầu thu đã thu được một tin nào đó.

Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Entropy có điều kiện (2/2)

Định nghĩa (Entropy có điều kiện)

Với một cặp nguồn rời rạc (X, Y) có xác suất phân bố hợp $p(x_k, y_l)$, , xác xuất phân bố có điều kiện $p(x_k|y_l)$, Entropy có điều kiện $H(Y|X)$ được cho bởi công thức:

$$\begin{aligned} H(Y|X) &\triangleq E[H(Y|X=x_k)]_{x_k \in X} = \sum_{x_k \in X} p(x_k) H(Y|X=x_k) \\ &= -\sum_{x_k \in X} p(x_k) \sum_{y_l \in Y} p(y_l|x_k) \log(p(y_l|x_k)) \\ &= -\sum_{k=1}^{N} \sum_{l=1}^{M} p(x_k, y_l) \log(p(y_l|x_k)) \\ &= E[-\log(p(Y|X))]_{p(x_k,y_l)} \end{aligned}$$

$H(Y|X)$: lượng tin riêng trung bình chứa trong mỗi tin ở đầu thu khi đầu phát đã phát đi một tin nào đó.

Notes

C2: Lý thuyết thông tin thống kê

Nội dung chính

Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Entropy tương đối và lượng thông tin tương hỗ giữa các nguồn: Entropy tương đối

Định nghĩa (Entropy tương đối - Relative Entropy)

Entropy tương đối, còn gọi là khoảng cách Kullback Leibler giữa hai phân bố rời rạc $p(x_k)$ và $q(x_k)$ của một nguồn rời rạc X được xác định bởi:

$$D(p||q) \triangleq \sum_{k=1}^{N} p(x_k) \log\left(\frac{p(x_k)}{q(x_k)}\right)$$

- Quy ước: $0 \log(\frac{0}{q}) = 0$; $p \log(\frac{p}{0}) = \infty$
- Tính chất:
  - $D(p||q) \geq 0$, $D(p||q) = 0$ nếu và chỉ nếu $p(x_k) = q(x_k)$
  - Tổng quát $D(p||q) \neq D(q||p)$
  - Không thỏa mãn $D(p||q) + D(q||r) \geq D(p||r) \Rightarrow$ không phải khoảng cách thông thường.

Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Entropy tương đối và lượng thông tin tương hỗ giữa các nguồn: Lượng thông tin tương hỗ giữa các nguồn

Định nghĩa (Lượng thông tin tương hỗ - Mutual Information)

Cho hai nguồn rời rạc X, Y có các xác suất phân bố hợp, phân bố riêng, và phân bố có điều kiện lần lượt là $p(x_k, y_k)$, $p_X(x_k) = p(x_k)$, $p_Y(y_l) = p(y_l)$, và $p(x_k|y_l)$. Lượng thông tin tương hỗ, còn gọi là lượng thông tin chéo trung bình của hai nguồn được xác định bởi:

$$I(X;Y) \triangleq E[I(x_k;y_l)] = \sum_{k=1}^{N}\sum_{l=1}^{M} p(x_k,y_l)\log\left(\frac{p(x_k|y_l)}{p(x_k)}\right)$$
$$= \sum_{k=1}^{N}\sum_{l=1}^{M} p(x_k,y_l)\log\left(\frac{p(x_k,y_l)}{p(x_k)p(y_l)}\right)$$
$$= D(p(x_k,y_l)||p(x_k)p(y_l))$$

- $I(X;Y)$: lượng thông tin mà X cho biết về Y cũng như lượng thông tin Y cho biết về X.

Notes

C2: Lý thuyết thông tin thống kê

Nội dung chính

Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Tính chất và các mối quan hệ giữa các đại lượng: Các tính chất của Entropy, ví dụ minh họa

- $H(X) \geq 0$, $H(X) = 0$ khi và chỉ khi $p(x_k) = 1$ và $p(x_r) = 0$ $(\forall r \neq k)$
- $H(X) \leq \log|X| = \log(N)$, $H(X) = \log(N)$ khi và chỉ khi các x_k có phân bố xác suất đồng đều, $p(x_k) = 1/N$ $\forall k$
- $H(X)$ là một hàm chỉ phụ thuộc vào đặc tính thống kê của nguồn
- $H_b(X) = (\log_b(a))H_a(X)$, $H_a(X)$: entropy được tính với cơ số a; Quy ước: $H(X)$ cơ số 2.

Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Tính chất và các mối quan hệ giữa các đại lượng: Các tính chất của Entropy có điều kiện, Entropy hợp (1/2)

$0 \leq H(X|Y) \leq H(X)$; $0 \leq H(Y|X) \leq H(Y)$

- Đạt đẳng thức phía phải khi và chỉ khi X và Y là độc lập: kênh bị đứt.
- Đạt đẳng thức phía trái khi và chỉ khi kênh hoàn hảo.

Nếu X và Y độc lập

- $H(X|Y = y_l) = H(X)$; $H(X|Y) = H(X)$.
- $H(Y|X = x_k) = H(Y)$; $H(Y|X) = H(Y)$.

Trường hợp tổng quát $H(X|Y) \neq H(Y|X)$.

$$H(X,Y) = H(Y,X) = H(X) + H(Y|X) = H(Y) + H(X|Y)$$

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, X_{i-2}, \ldots, X_1)$$

Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Tính chất và các mối quan hệ giữa các đại lượng: Các tính chất của Entropy có điều kiện, Entropy hợp (2/2)

$$H(X,Y) \le H(X) + H(Y)$$

$$H(X_1, X_2, \ldots, X_n) \le \sum_{i=1}^{n} H(X_i)$$

- Xảy ra đẳng thức khi và chỉ khi X và Y độc lập: kênh bị đứt.

$H(X,Y|Z) = H(X|Z) + H(Y|X,Z)$.

Cho nguồn rời rạc X, $g()$ là một hàm mô tả quan hệ toán học xác định, khi đó:

- $H(g(X)|X) = 0$
- $H(X|g(X)) \ge 0$
- $H(X) \ge H(g(X))$
  - Xảy ra đẳng thức khi và chỉ khi $g()$ là quan hệ toán học $1-1$.

Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Tính chất và các mối quan hệ giữa các đại lượng: Các tính chất của lượng tin tương hỗ

$0 \le I(X;Y) \le H(X)$, $0 \le I(X;Y) \le H(Y)$

- Xảy ra đẳng thức bên phải khi và chỉ khi X và Y độc lập
- Xảy ra đẳng thức bên trái khi và chỉ khi kênh lý tưởng không nhiễu

$I(X;Y) = I(Y;X)$

- Lượng thông tin mà X cho biết về Y cũng bằng lượng thông tin mà Y cho biết về X.

$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$

- $I(X;Y)$: lượng giảm độ bất định trung bình của X do việc biết Y.

$I(X;Y) = H(X) + H(Y) - H(X,Y)$

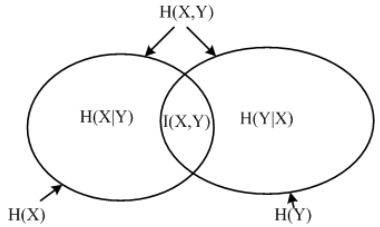
$I(X;X) = H(X)$

- $H(X)$: lượng thông tin riêng trung bình của X.

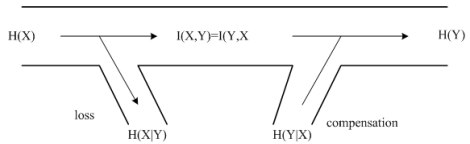
Notes

Entropy và các đại lượng liên quan của nguồn rời rạc

Tính chất và các mối quan hệ giữa các đại lượng: Biểu diễn mối liên hệ giữa các đại lượng



(a) Biểu đồ Venn



(b) Sơ đồ kênh

Notes

C2: Lý thuyết thông tin thống kê

Nội dung chính

Notes

Phần II

Lý thuyết thông tin thống kê cho nguồn liên tục

Notes

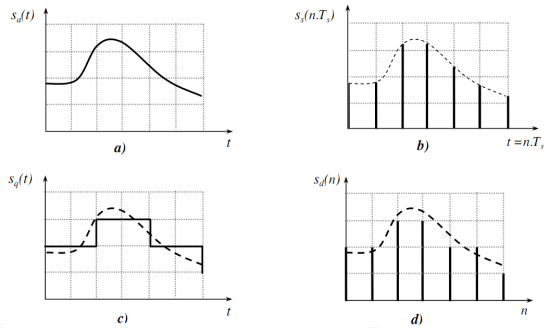
C2: Lý thuyết thông tin thống kê

Nội dung chính

Notes

Tín hiệu liên tục, nguồn liên tục

Tín hiệu liên tục: Minh họa đồ thị các loại tín hiệu



Tín hiệu:

- Biểu diễn: hàm toán học của các biến độc lập
- Đặc trưng tín hiệu liên tục: Công suất phổ trung bình, bề rộng phổ

Notes

Tín hiệu liên tục, nguồn liên tục

Nguồn liên tục

Nguồn liên tục

Nguồn tin X phát ra các tin x có giá trị liên tục trong khoảng $x_{min} \div x_{max}$ với hàm mật độ phân bố xác suất $f(x)$

Mô hình toán học nguồn liên tục

- Biến ngẫu nhiên liên tục X với hàm mật độ phân bố xác suất $f(x)$

Notes

Nội dung chính

Notes

Entropy vi phân

Định nghĩa (Entropy vi phân - Differential Entropy)

Entropy vi phân của một nguồn liên tục X có hàm mật độ phân bố xác suất $f(x)$ được xác định bởi:

$$h(X) \triangleq -\int_S f(x)\log(f(x))dx$$

trong đó, S là miền xác định dương (support set: tập trên đó $f(x) \geq 0$) của X.

- $h(X)$ mặc định chỉ xem xét trên điều kiện các hàm liên tục, xác định và khả tích.
- $h(X) = h(f)$

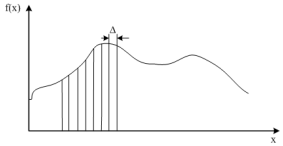
Ví dụ

Cho X là một nguồn liên tục có hàm mật đố phân bố xác suất đều (uniform distribution) trong đoạn $[a, b]$. Tính $h(X)$.

Notes

Entropy vi phân

Mối quan hệ giữa Entropy vi phân và Entropy rời rạc



- $X \to X^\Delta = \{x_i\}$ $(i\Delta \le X \le (i+1)\Delta)$.
- $p(X^\Delta = x_i) = p(x_i) = f(x_i)\Delta = \int_{i\Delta}^{(i+1)\Delta} f(x)dx$
- $\to H(X^\Delta) = -\sum_{-\infty}^{\infty} f(x_i)\Delta \log(f(x_i)\Delta) =$ $-\sum_{-\infty}^{\infty} \Delta f(x_i)\log(f(x_i)) + \log(1/\Delta)$

Định lý

Một nguồn liên tục X với hàm mật độ phân bố xác suất $f(x)$ khả tích theo tiêu chuẩn Riemann thì:

$$H(X^\Delta) + \log(\Delta) \to h(X) \text{ khi } \Delta \to 0$$

- Entropy của một nguồn rời rạc thu được từ nguồn liên tục X bằng phép lượng tử hóa sử dụng n bít có giá trị xấp xỉ bằng $h(X) + n$

Notes

Entropy vi phân

Minh họa Entropy của nguồn liên tục

$\lim_{\Delta\to 0} \log(\frac{1}{\Delta}) \to \infty \Rightarrow H(X)$ lớn vô hạn.

Ví dụ

Xét việc truyền thông tin từ nguồn liên tục X đến nguồn Y bằng dây dẫn lý tưởng (không tổn hao, không nhiễu). Tín hiệu phát $x(t)$ nhận các giá trị liên tục trong khoảng $[0, 1]$ (V). Ở đầu thu Y ta đặt một vôlt kế lý tưởng (tạp âm nội bằng 0, $Z_V = \infty$). Khi đó việc thu tín hiệu thỏa mãn $y(t) = x(t)$. Xem xét việc lượng tử hóa, và tính toán $H(X)$.

Lượng tử đều:

- 10 mức $\Delta = 0,1$: $X^\Delta = \{x_i\}$ $(i = 1,\overline{10}) \Rightarrow H(X^\Delta) = \log(10)$.
- $\Delta = 0,01 \Rightarrow H(X^\Delta) = \log(100)$.
- $\Delta \to 0 \Rightarrow H(X^\Delta) \to H(X) \to \infty$.

Notes

Entropy vi phân

Một số tính chất

- $h(X)$ có thể âm, dương.
- $h(X)$ có giá trị hữu hạn.
- Với một hằng số c: $h(X + c) = h(X)$
- Với một hằng số $c \neq 0$: $h(cX) = h(X) + \log(|c|)$
  - $h(X)$ phụ thuộc vào thang tỷ lệ (đơn vị đo)

Định lý

Trong số những quá trình ngẫu nhiên (tín hiệu) có cùng công suất trung bình $P_x = \sigma^2$, quá trình (tín hiệu) có hàm mật độ phân bố chuẩn (phân bố Gausse) sẽ cho Entropy vi phân lớn nhất. Nói cách khác

$$h(X) \leq \log(\sqrt{2\pi e P_x})$$

- Trong số các tín hiệu nhiễu (tạp âm) có cùng công suất trung bình, tín hiệu nhiễu Gausse có tác hại lớn nhất với việc truyền tin.

Notes

C2: Lý thuyết thông tin thống kê

Nội dung chính

Notes

Entropy vi phân hợp, Entropy vi phân có điều kiện, Lượng tin tương hỗ của các nguồn liên tục

Entropy vi phân hợp, Entropy vi phân có điều kiện

Định nghĩa (Entropy vi phân hợp)

Entropy vi phân hợp của cặp nguồn liên tục (X, Y) với hàm mật độ phân bố hợp (phân bố đồng thời) $f(x, y)$, được định nghĩa:

$$h(X, Y) \triangleq -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\log(f(x, y))dxdy$$

Định nghĩa (Entropy vi phân có điều kiện)

Các Entropy vi phân có điều kiện của cặp nguồn liên tục (X, Y) với hàm mật độ phân bố hợp (phân bố đồng thời) $f(x, y)$, được định nghĩa:

$$h(X|Y) \triangleq -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\log(f(x|y))dxdy$$

$$h(Y|X) \triangleq -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\log(f(y|x))dxdy$$

Notes

Lượng thông tin tương hỗ của các nguồn liên tục

Định nghĩa (Entropy vi phân tương đối)

Xét một nguồn liên tục X, với nguồn X giả sử có hai phân bố $f(x)$ và $g(x)$. Entropy vi phân tương đối hay còn gọi là khoảng cách Kullback Leibler được tính bằng công thức:

$$D(f(x)||g(x)) \triangleq \int_S f(x)\log\left(\frac{f(x)}{g(x)}\right)dx$$

- $D(f||g) < \infty$ iff miền xác định (support set) của $f()$ chứa miền của $g()$.
- Quy ước $0\log\frac{0}{0} = 0$

Định nghĩa (Lượng thông tin tương hỗ)

Lượng thông tin tương hỗ $I(X; Y)$ giữa hai nguồn liên tục X và Y có xác suất phân bố hợp $f(x, y)$ được xác định bởi công thức:

$$I(X; Y) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\log\left(\frac{f(x, y)}{f(x)f(y)}\right)dxdy$$

Notes

Entropy vi phân hợp, Entropy vi phân có điều kiện, Lượng tin tương hỗ của các nguồn liên tục

Một số tính chất (1)

$h(X, Y) = h(X) + h(Y|X) = h(Y) + h(X|Y)$

$h(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} h(X_i|X_{i-1}, \ldots, X_1)$

$h(X|Y) \leq h(X)$; $h(Y|X) \leq h(Y)$

- Xảy ra đẳng thức khi và chỉ khi X và Y độc lập nhau.

$h(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} h(X_i)$

$D(f(x)|g(x)) \geq 0$

- Xảy ra đẳng thức iff $f() = g()$ trên gần toàn miền xác định.

Notes

Entropy vi phân hợp, Entropy vi phân có điều kiện, Lượng tin tương hỗ của các nguồn liên tục

Một số tính chất (2)

$I(X; Y) = D(f(x, y)|f(x)f(y))$

$I(X; Y) \geq 0$

- Xảy ra đẳng thức iff X và Y độc lập nhau.

$I(X; Y) = I(Y; X)$

$I(X; Y) = h(X) - h(X|Y) = h(Y) - h(Y|X)$

$I(X^{\Delta}; Y^{\Delta}) \approx I(X; Y)$

- $I(X; Y)$ là giới hạn của lượng thông tin tương hỗ giữa các nguồn rời rạc hóa (lượng tử hóa) tương ứng.

Notes

Kết thúc bài học

Notes

Notes

C2(cont.): Dung lượng kênh

Lý thuyết thông tin

Biên soạn: Phạm Văn Sự

Bộ môn Xử lý tín hiệu và Truyền thông
Khoa Kỹ thuật Điện tử I
Học viện Công nghệ Bưu chính Viễn thông

ver. 22a

Notes

Mục tiêu của bài học

- Trang bị một số khái niệm cơ bản về kênh truyền
- Cách xác định dung lượng của kênh
- Mối quan hệ giữa tốc độ dữ liệu và dung lượng kênh để đảm bảo truyền tin tin cậy

Notes

Các câu hỏi cần trả lời

- Tốc độ dữ liệu đầu vào kênh được đánh giá thế nào?
- Thế nào là kênh giãn tin? kênh nén tin? kênh thông thường?
- Thế nào là kênh rời rạc không nhớ? Kênh đối xứng? Kênh đồng nhất?
- Lượng tin trung bình truyền qua kênh rời rạc không nhớ?
- Dung lượng của một kênh rời rạc không nhớ xác định bằng công thức nào? Có tính chất gì? Cách xác định cho các bài toán cụ thể?
- Thế nào là một kênh AWGN? Mô hình?
- Lượng tin trung bình truyền qua kênh AWGN?
- Dung lượng của một kênh AWGN được xác định thế nào? Dung lượng một kênh AWGN có băng thông hữu hạn?
- Định lý mã hóa thứ hai của Shannon?

Notes

C2(cont.): Dung lượng kênh

Nội dung chính

Notes

C2(cont.): Dung lượng kênh

Nội dung chính

Notes

Tốc độ phát, khả năng phát của nguồn rời rạc

Tốc độ phát, khả năng phát của nguồn rời rạc

Định nghĩa (Tốc độ phát của nguồn rời rạc)

Tốc độ phát của một nguồn rời rạc được định nghĩa $v_n = \frac{1}{T_n}$

- T_n: độ rộng trung bình của mỗi xung phát.
- v_n: số xung phát trong một đơn vị thời gian, v_n: đơn vị [baud]

Định nghĩa (Khả năng phát của nguồn rời rạc)

Một nguồn rời rạc X có tốc độ phát $v_n = \frac{1}{T_n}$*, khi đó khả năng phát của nguồn được xác định:*

$$H'(X) = v_n H(X) = \frac{H(X)}{T_n}$$

- $H'(X)$: lượng thông tin trung bình do nguồn phát ra trong một đơn vị thời gian, Đơn vị [bit/s]
- $H'(X)_{max} = v_n \log(N) = \log(N)/T_n$

Notes

Tốc độ phát, khả năng phát của nguồn rời rạc

Độ dư thừa của nguồn

Định nghĩa (Độ dư thừa của nguồn rời rạc)

Với một nguồn rời rạc X, một phép xử lý thông tin đạt được $H(X)$, khi đó độ dư thừa của nguồn được định nghĩa là:

$$D = \frac{H(X)_{max} - H(X)}{H(X)_{max}} = 1 - \frac{H(X)}{H(X)_{max}} = 1 - \mu$$

- $\mu = \frac{H(X)}{H(X)_{max}}$: là tỷ số nén tin.
- D đặc trưng cho hiệu suất, khả năng chống nhiễu và mật độ của tin
  - D lớn $\Rightarrow$ hiệu suất thấp, khả năng chống nhiễu cao.

Notes

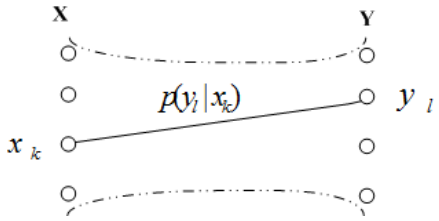
C2(cont.): Dung lượng kênh

Nội dung chính

Notes

Kênh rời rạc và một số khái niệm

Đặc trưng của kênh rời rạc



Một kênh rời rạc hoàn toàn có thể đặc trưng bởi 3 tham số:

- Trường tin lối vào X (input), trường tin lối ra Y (output).
- Xác suất chuyển tin lối vào x_k thành tin lối ra y_l: $p(y_l|x_k)$.
- Tốc độ truyền tin của kênh v_k hay thời gian trung bình để truyền một dấu tin $T_k = \frac{1}{v_k}$.

Notes

Kênh rời rạc và một số khái niệm

Đặc trưng của kênh rời rạc - Một số khái niệm

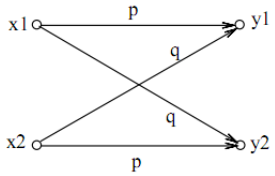
Định nghĩa (Kênh đồng nhất)

Xét một kênh rời rạc có xác suất chuyển $p(y_l|x_k)$.

- *Nếu $p(y_l|x_k)$ không phụ thuộc vào thời gian t thì kênh được gọi là kênh đồng nhất; ngược lại gọi là kênh không đồng nhất.*

Định nghĩa (Kênh đối xứng)

Xét một kênh rời rạc có xác suất chuyển $p(y_l|x_k)$. Nếu $p(y_l|x_k) = p = const\ \forall k, l$, $k \neq l$ và $p(y_l|x_k) = q = const\ \forall k = l$ thì kênh được gọi là đối xứng.



Hình: Mô hình kênh nhị phân đối xứng (BSC)

Notes

Kênh rời rạc và một số khái niệm

Đặc trưng của kênh rời rạc - Một số khái niệm (cont.)

Định nghĩa (Kênh không có nhớ)

Nếu $p(y_l|x_k)$ không phụ thuộc vào các tin (kí hiệu) phát/nhận trước đó thì kênh được gọi là kênh không có nhớ (memoryless):

$$p(y_l|x_k, x_{k-1}, \ldots, x_1, y_{l-1}, \ldots, y_1) = p(y_l|x_k)$$

- Nếu y_k tương ứng với tin phát $x_k \Rightarrow$

  $p(y_1, y_2, \ldots, y_n|x_1, x_2, \ldots, x_n) = \prod_{k=1}^{n} p(y_k|x_k)$

Biểu diễn kênh:

- Giản đồ chuyển trên đó nhãn các đường chuyển là các $p(y_l|x_k)$
- Các ma trận xác suất chuyển $\mathrm{P} = [p_{kl}]$ với $p_{kl} = p(y_l|x_k)$

$$\mathrm{P} = \begin{bmatrix} p(y_1|x_1) & p(y_2|x_1) & \ldots & p(y_M|x_1) \\ p(y_1|x_2) & p(y_2|x_2) & \ldots & p(y_M|x_2) \\ \vdots & \vdots & \ddots & \vdots \\ p(y_1|x_N) & p(y_2|x_N) & \ldots & p(y_M|x_N) \end{bmatrix}$$

Notes

Kênh rời rạc và một số khái niệm

Lượng thông tin truyền qua kênh trong một đơn vị thời gian

Định nghĩa

Một kênh rời rạc có lượng tin truyền qua $I(X;Y)$ với tốc độ truyền tin v_k thì lượng thông tin truyền qua kênh trong một đơn vị thời gian là:

$$I'(X;Y) = v_k I(X;Y) = \frac{I(X;Y)}{T_k}$$

- $T_k > T_n$: kênh giãn tin
- $T_k = T_n$: kênh thông thường
- $T_k < T_n$: kênh nén tin

Notes

C2(cont.): Dung lượng kênh

Nội dung chính

Notes

Dung lượng kênh rời rạc

Định nghĩa (Khả năng thông qua của kênh rời rạc)

Khả năng thông qua của kênh rời rạc là giá trị cực đại của lượng thông tin truyền qua kênh trong một đơn vị thời gian lấy theo mọi khả năng có thể của phân bố nguồn phát.

$$\begin{aligned} C' &= \max_{p(X)} I'(X;Y) = \max_{X} I'(X;Y) = v_k \max_{X} I(X;Y) \quad [bit/s] \\ &= v_k C \end{aligned}$$

- $C = \max_X I(X;Y)$: khả năng thông qua của kênh đối với mỗi dấu.
  - C: đơn vị [bít/lần truyền]
  - C: thường được sử dụng.

Tính chất:

- $C' \geq 0$, $C' = 0$ khi và chỉ khi X và Y hoàn toàn độc lập $\Rightarrow$ kênh bị đứt.
- $C' \leq v_k \log(N)$ (N là độ lớn của nguồn X)
- $C' \leq v_k \log(M)$ (M là độ lớn của nguồn Y)

Notes

C2(cont.): Dung lượng kênh

Nội dung chính

Notes

Định lý mã hóa thứ hai của Shannon

Định lý mã hóa thứ hai của Shannon

Định lý

Nếu khả năng phát $H'(X)$ của một nguồn rời rạc X nhỏ hơn khả năng thông qua của kênh ($H'(X) \leq C'$) thì tồn tại một phép mã hóa và giải mã sao cho việc truyền tin qua kênh có xác suất lỗi nhỏ tùy ý khi độ dài từ mã đủ lớn. Ngược lại thì không tồn tại một phép mã hóa nào như vậy.

Định lý

Nếu tốc độ dữ liệu cần truyền R truyền qua kênh có dung lượng C' thỏa mãn $R \leq C'$ thì tồn tại một phép mã hóa và giải mã sao cho việc truyền tin qua kênh có xác suất lỗi nhỏ tùy ý khi độ dài từ mã đủ lớn. Ngược lại thì không tồn tại một phép mã hóa nào như vậy.

- *Nhận xét*: Định lý chỉ ra tính tồn tại, không chỉ ra cách xây dựng

Notes

C2(cont.): Dung lượng kênh

Nội dung chính

Notes

C2(cont.): Dung lượng kênh

Nội dung chính

1. Dung lượng kênh rời rạc
   - Khả năng phát của nguồn rời rạc
   - Kênh rời rạc và một số khái niệm
   - Dung lượng kênh rời rạc
   - Định lý mã hóa thứ hai của Shannon
2. Dung lượng kênh Gausse nhiễu trắng cộng
   - Kênh Gausse nhiễu trắng cộng - AWGN
   - Dung lượng của kênh AWGN

Notes

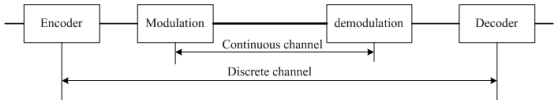
Kênh Gausse nhiễu trắng cộng - AWGN

Đặc trưng của kênh Gausse nhiễu cộng

Tham số đặc trưng của kênh liên tục:

- Trường dấu lối vào (input) và trường dấu lối ra (output).
- Hàm chuyển, hàm mật độ phân bố xác suất để thu được $y(t)$ khi đã phát $x(t)$: $f(y(t)|x(t))$
- Tốc độ truyền của kênh v_k

Kênh rời rạc chứa kênh liên tục:



Định lý

Khả năng thông qua của kênh liên tục không nhỏ hơn khả năng thông qua của kênh rời rạc chứa nó.

$$C_{\text{liên tục}} \geq C_{\text{rời rạc chứa liên tục}}$$

Notes

Kênh Gausse nhiễu trắng cộng - AWGN

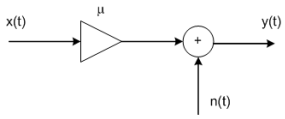
Mô hình của kênh Gausse nhiễu cộng

Định nghĩa (Kênh Gausse)

Kênh Gausse không đổi là một kênh liên tục có tập tin lối vào và tập tin lối ra liên hệ với nhau theo công thức:

$$y(t) = \mu x(t) + n(t)$$

trong đó: $\mu = const$; $n(t)$ là nhiễu cộng còn gọi là nhiễu trắng có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$.

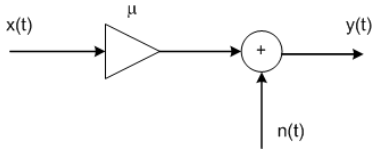


- $x(t) \sim X$, $y(t) \sim Y$, và $n(t) \sim N$: $N \sim \mathcal{N}(0, \sigma_n^2)$
- $x(t)$ và $n(t)$ độc lập nhau.
- $\rightarrow \sigma_y^2 = \mu^2\sigma_x^2 + \sigma_n^2$ hay tương đương $P_y = \mu^2 P_x + P_n$

Notes

Kênh Gausse nhiễu trắng cộng - AWGN

Lượng thông tin tương hỗ qua kênh AWGN



- $y(t) = \mu x(t) + n(t)$
- Giả sử: $n(t) \sim \mathcal{N}(0, \sigma_n^2)$, $x(t)$ và $y(t)$ cũng có phân bố chuẩn

- $I(X;Y) = h(Y) - h(Y|X)$
- $h(Y) = log\sqrt{2\pi e P_y}$ trong đó $P_y = \mu^2 P_x + P_n$
- $h(Y|X) = -\int\int f(x,y) log(f(y|x)) dxdy$
  - $Pr\{y \in dy|x\} = Pr\{n \in dn\} \rightarrow f(y|x)dy = f(n)dn \rightarrow f(y|x) = f(n)\frac{dn}{dy} = f(n)\frac{1}{\frac{dy}{dn}} = f(n)$
  - $\Rightarrow h(Y|X) = log\sqrt{2\pi e P_n}$
- $\Rightarrow I(X;Y) = \frac{1}{2} log\left(1 + \frac{P_x}{P_n}\right)$

Notes

C2(cont.): Dung lượng kênh

Nội dung chính

Notes

Dung lượng kênh AWGN

Định nghĩa (Dung lượng của kênh liên tục)

Khả năng thông qua của kênh liên tục, còn gọi là dung lượng kênh liên tục, là giá trị cực đại của lượng thông tin truyền qua kênh trong một đơn vị thời gian lấy theo mọi khả năng có thể của phân bố nguồn phát trong đó kể đến giới hạn công suất phát.

$$C' = v_k \max_{f(x):E\{x^2(t)\}\le P} I(X;Y) = v_k \max_{X:E\{x^2(t)\}\le P} I(X;Y)$$

$$C = \max_{f(x):E\{x^2(t)\}\le P} I(X;Y) = \max_{X:E\{x^2(t)\}\le P} I(X;Y)$$

- $v_k = \frac{1}{\Delta t}$ với Δt là thời gian rời rạc hóa.

Định lý (Dung lượng của kênh Gausse nhiễu cộng)

Kênh AWGN với giới hạn công suất phát P và công suất nhiễu N có dung lượng:

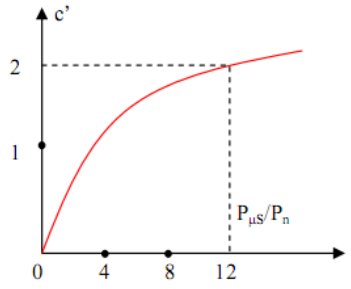
$$C' = \frac{v_k}{2}\log\left(1+\frac{\mu^2 P}{P_n}\right) \qquad C = \frac{1}{2}\log\left(1+\frac{\mu^2 P}{P_n}\right)$$

Notes

Dung lượng kênh AWGN

Một số nhận xét

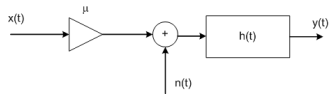
- $\mu^2 P/P_n = S/N$ gọi là tỷ số công suất trung bình của tín hiệu trên tạp âm (SNR)
- $S/N \to 0 \Rightarrow C' \to 0$: S/N rất bé thì có thể coi như kênh bị đứt.
- Để tăng C' thì cần tăng S/N, tuy nhiên việc tăng này bị giới hạn do C' rơi vào tình trạng bão hòa.



Notes

Dung lượng kênh AWGN

Kênh có băng thông hạn chế



- $y(t) = (\mu x(t) + n(t)) * h(t)$
- Kênh AWGN với mật độ phổ công suất nhiễu hai phía $N_0/2$ [W/Hz].
- $h(t)$: đáp ứng xung của một mạch lọc thông dải lý tưởng có băng tần W [Hz]. $\rightarrow$ Tốc độ lấy mẫu $\geq \frac{1}{2W}$

Định lý (Dung lượng của kênh AWGN băng tần hữu hạn)

Dung lượng của kênh AWGN với băng tần hữu hạn W và giới hạn công suất phát P_x có nhiễu với mật độ phổ công suất hai phía $N_0/2$ được xác định:

$$C' = W \log\left(1 + \frac{\mu^2 P_x}{N_0 W}\right) [bps]$$

Notes

Dung lượng kênh AWGN

Kênh có băng tần hạn chế - Khảo sát ảnh hưởng của băng tần

- $W \longrightarrow 0 \Rightarrow C' \longrightarrow 0$
- Nếu $W \uparrow$, thì $C' \uparrow$.
  - **Chú ý**: $W \uparrow \rightarrow P_n = WN_0 \uparrow \rightarrow SNR \downarrow$
- $W \rightarrow \infty$, $C \rightarrow C'_\infty = \frac{\mu^2 P_x}{N_0} \log_2 e < \infty$.
  - Thông tin vũ trụ thường với băng tần rất rộng.
- $0 \leq C' \leq C'_\infty$
  - **Chú ý**: Tạp âm nhiệt luôn tồn tại.

Định lý (Định lý mã hóa thứ hai của Shannon)

Các nguồn tin rời rạc có thể mã hóa và truyền theo kênh liên tục với xác suất sai bé tùy ý khi giải mã các tín hiệu nhận được nếu khả năng phát của nguồn nhỏ hơn khả năng thông qua của kênh. Ngược lại, không thể thực hiện được phép mã hóa và giải mã với sai số bé tùy ý.

Chú ý: $\lim_{x \to 0}(1 + x)^{1/x} = e$

Notes

Kết thúc phần dung lượng kênh

Notes

Notes

C3: Mã hóa nguồn - Nén dữ liệu

Lý thuyết thông tin

Biên soạn: Phạm Văn Sự

Bộ môn Xử lý tín hiệu và Truyền thông
Khoa Kỹ thuật Điện tử I
Học viện Công nghệ Bưu chính Viễn thông

ver. 22a

Notes

Mục tiêu của bài học

- Trang bị một số khái niệm cơ bản về mã hóa, mã hóa nguồn
- Mã hóa tối ưu cho nguồn
- Một số thuật toán mã hóa nguồn phổ biến

Notes

Các câu hỏi cần trả lời

- Mã hóa là gì? Bộ mã là gì? Các thông số cơ bản của một bộ mã?
- Thế nào là mã không suy biến? mã có khả năng giải mã duy nhất? mã có tính prefix?
- Bài toán mã hóa tối ưu? Cách để xây dựng được bộ mã tối ưu?
- Định lý mã hóa thứ nhất của Shannon?
- Mã hóa khối tin có ưu điểm gì so với mã hóa đơn lẻ từng tin?
- Những điểm cơ bản nhất về các thuật toán mã hóa: Shannon, Shannon-Fano, Shannon-Fano-Elias, mã hóa số học, LZW?
- Cách thức thực hiện mã hóa và giải mã cho các phương thức mã hóa : Shannon, Shannon-Fano, Shannon-Fano-Elias, mã hóa số học, LZW?

Notes

C3: Mã hóa nguồn - Nén dữ liệu

Nội dung chính

Notes

Tổng quan về mã hóa nguồn

Mục tiêu và phân loại

Mục tiêu của mã hóa nguồn

Thực hiện tìm kiếm các phương thức biểu diễn dữ liệu nhỏ gọn nhất có thể

Nguyên lý của mã hóa nguồn

Loại bỏ các thông tin dư thừa hoặc các thông tin dư thừa và các thông tin không cần thiết.

- Theo quan điểm bảo toàn thông tin:
  - Nén không tổn hao (lossless data compression)
  - Nén có tổn hao (lossy data compression)
- Theo đặc tính thay đổi:
  - Mã thích nghi (adaptive)
  - Mã không thích nghi (nonadaptive)
- Theo phương pháp:
  - RLE (run length encoding)
  - Mã hóa thống kê
  - Mã hóa từ điển
  - Mã hóa chuyển đổi
- Theo mô hình n-user:
  - Tập trung
  - Phân tán

Notes

C3: Mã hóa nguồn - Nén dữ liệu

Nội dung chính

Notes

Các định nghĩa và khái niệm cơ bản mã hóa

Mã hóa

Cho nguồn rời rạc X với các tin x_k có xác suất phân bố $p(x_k)$. Một bộ dấu (chữ mã) M với các dấu (chữ mã) $\{m_1, m_2, \ldots, m_q\}$.

Định nghĩa (Mã hóa)

Mã hóa là một phép ánh xạ $1-1$ từ tập các tin rời rạc x_k lên tập các từ mã là tổ hợp có thể của các dấu (các chữ mã) m_k

$$f : x_k \longmapsto m_k^{l_k}$$

- l_k là độ dài từ mã thứ k: số dấu mã tạo thành từ mã $m_k^{l_k}$.
- $m_k^{l_k}$ gọi là từ mã.
  - $m_k^{l_k}$ thường là các phần tử của một cấu trúc đại số
  - Các dấu mã thường được chọn từ một trường F nào đó
- Bộ mã: tập hợp các từ mã, là sản phẩm của phép mã hóa.

Notes

Các định nghĩa và khái niệm cơ bản mã hóa

Các thông số cơ bản của bộ mã

- Độ dài từ mã: l_k là độ dài từ mã thứ k; $l_k = const\ \forall k$ gọi là mã đều, ngược lại gọi là mã không đều.
- Độ dài trung bình: là trung bình thống kê của độ dài các từ mã: $\bar{l} = \sum_{k=1}^{N} p(x_k) l_k$
- Cơ số mã: số các dấu (chữ mã) khác nhau được sử dụng trong bộ mã.
- Bộ mã mà tất cả các tổ hợp dấu mã là từ mã của tập tin tương ứng gọi là bộ mã đầy, ngược lại gọi là mã không đầy (mã vơi).
- Tính hiệu quả của phép mã hóa: $\eta = \frac{\bar{l}_{min}}{\bar{l}} = \frac{H(X)}{\bar{l}} \rightarrow \eta \leq 1$. Bộ mã hiệu quả khi $\eta \rightarrow 1$.
- Độ chậm giải mã: là số dấu (chữ mã) nhận được cần thiết trước khi có thể thực hiện được việc giải mã.
- Phương sai độ dài trung bình của bộ mã $\sigma_l^2 = \sum_{k=1}^{N} p(x_k)(l_k - \bar{l})^2$

Notes

Các định nghĩa và khái niệm cơ bản mã hóa

Khái niệm các bộ mã (1)

Định nghĩa (Mã không suy biến (không dị thường))

Một bộ mã được gọi là không suy biến (non-singular) nếu mọi tin x_k của nguồn X ánh xạ thành các từ mã khác nhau của bộ mã.

$$x_k \neq x_l \Rightarrow m_k^{l_k} \neq m_l^{l_l}$$

- Đảm bảo cho việc mô tả không bị nhập nhằng giữa các tin sau khi mã hóa

Định nghĩa (Từ mã mở rộng)

Một từ mã mở rộng là việc ánh xạ một chuỗi hữu hạn các tin thành các từ mã liên tiếp nhau.

$$x_1 x_2 \cdots \longmapsto m_1^{l_1} m_2^{l_2} \ldots$$

Định nghĩa (Bộ mã có khả năng giải mã một cách duy nhất)

Một bộ mã được gọi là bộ mã có khả năng giải mã được một cách duy nhất nếu từ mã mở rộng của nó là một từ mã không suy biến.

Notes

Các định nghĩa và khái niệm cơ bản mã hóa

Khái niệm các bộ mã (2)

Định nghĩa (Bộ mã có tính prefix)

Một bộ mã được gọi là bộ mã có tính prefix hay còn gọi mã có khả năng giải mã tức thời nếu không có bất cứ từ mã nào là phần tiền tố (prefix) của một từ mã khác trong bộ mã.

- Một bộ mã prefix là bộ mã có khả năng tự phân tách được.

Định lý (Bất đẳng thức Kraft)

Với bất cứ bộ mã prefix nào trên tập dấu (chữ mã) M có kích thước (cơ số) q thì tập độ dài các từ mã có thể $l_1, l_2, \ldots, l_N$ phải thỏa mãn bất đẳng thức:

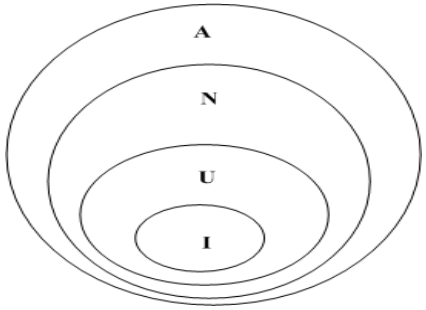
$$\sum_{k=1}^{N} q^{-l_k} \leq 1$$

Ngược lại, với một tập các độ dài từ mã cho trước thỏa mãn bất đẳng thức này thì tồn tại một bộ mã prefix nhận tập độ dài này làm độ dài các từ mã.

Notes

Các định nghĩa và khái niệm cơ bản mã hóa

Lược đồ Venn biểu diễn các bộ mã



Hình: Phân loại các lớp các mã (I) Mã giải mã tức thì (U) Mã có khả năng giải mã duy nhất (N) Mã không suy biến (A) Tất cả các mã

- Một bộ mã có khả năng giải mã một cách duy nhất chưa chắc là một bộ mã có tính prefix.

Notes

C3: Mã hóa nguồn - Nén dữ liệu

Nội dung chính

Notes

Nguyên tắc mã hóa tối ưu

Ví dụ, nguyên tắc mã hóa tối ưu

Ví dụ

Giả sử có bộ mã $\mathfrak{C} = \{0, 10, 110, 111\}$. Cho một đoạn văn bản sau: "aaaaabbbccd". Thực hiện việc mã hóa theo các phương án sau:

1. *Phương án 1 $a \leftrightarrow 111$, $b \leftrightarrow 110$, $c \leftrightarrow 10$ và $d \leftrightarrow 0$*
2. *Phương án 2 $d \leftrightarrow 111$, $c \leftrightarrow 110$, $b \leftrightarrow 10$ và $a \leftrightarrow 0$*

Tìm biểu diễn tương ứng của đoạn văn bản và so sánh các bản mã thu được.

Nguyên tắc

Gán các từ mã có độ dài ngắn cho các tin có xác suất xuất hiện lớn, và các từ mã có độ dài lớn cho các tin có xác suất xuất hiện nhỏ.

Notes

Nguyên tắc mã hóa tối ưu

Mã hóa tối ưu, Bài toán mã hóa tối ưu

Định nghĩa (Phép mã hóa tối ưu)

Một phép mã hóa được gọi là tiết kiệm (hay còn gọi là tối ưu) nếu nó đạt được độ dài trung bình từ mã cực tiểu $\bar{l}_{min}$

Bài toán mã hóa tối ưu

$$\min \bar{l} = \sum_k p(x_k) l_k$$

$$\text{sao cho} \sum_{k=1}^{N} q^{-l_k} \leq 1$$

$\Rightarrow l_k^* = -\log_q(p(x_k))$. Trường hợp tổng quát $l_k^* \notin Z^+$

Notes

Nguyên tắc mã hóa tối ưu

Đánh giá độ dài trung bình của mã tối ưu

Định lý

Độ dài trung bình từ mã $\bar{l}$ của bất cứ bộ mã có khả năng giải mã tức thì cơ số q nào biểu diễn một nguồn rời rạc X cũng lớn hơn hoặc bằng với entropy $H_q(X)$ của nguồn, nói cách khác:

$$\bar{l} \geq H_q(X)$$

xảy ra đẳng thức khi và chỉ khi $q^{-l_k} = p(x_k)$

- Phân bố thỏa mãn đẳng thức trên gọi là q-adic

Định lý

Gọi tập $l_1^*, l_2^*, ..., l_N^*$ là tập các độ dài từ mã tối ưu của phép mã hóa cơ số q cho nguồn rời rạc có phân bố p trên tập dấu mã M. Khi đó độ dài trung bình từ mã của bộ mã tối ưu $\bar{l}^*$ thỏa mãn bất đẳng thức kẹp:

$$H_q(X) \leq \bar{l}^* < H_q(X) + 1$$

Notes

Nguyên tắc mã hóa tối ưu

Mã khối dữ liệu

- Dãy n ký hiệu (tin) từ nguồn rời rạc X, mỗi tin x_k được lấy với xác suất phân bố độc lập tương đồng (i.i.d) $p(x_k)$.
- Gọi $l(x_1, x_2, \ldots, x_n)$ là độ dài từ mã tương ứng với dãy $(x_1, x_2, \ldots, x_n)$.
- Định nghĩa L_n là độ dài trung bình từ mã với mỗi ký hiệu, nói cách khác:

$$L_n = \frac{1}{n} \sum p(x_1, x_2, \ldots, x_n) l(x_1, x_2, \ldots, x_n) = \frac{1}{n} E[l(X_1, X_2, \ldots, X_n)]$$

Định lý

Độ dài trung bình từ mã với mỗi ký hiệu khi thực hiện mã hóa khối đồng thời thỏa mãn bất đẳng thức

$$H(X) \leq L_n < H(X) + \frac{1}{n}$$

- $n \longrightarrow \infty \Rightarrow L_n \longrightarrow H(X)$

Notes

Nguyên tắc mã hóa tối ưu

Mã hóa với đặc trưng thống kê xấp xỉ

Định lý

Độ dài trung bình bộ mã biểu diễn một nguồn có hàm mật độ phân bố $p(x)$ với các độ dài từ mã được sử dụng $l_k = \lceil \log \frac{1}{q(x_k)} \rceil$ thỏa mãn

$$H(p) + D(p||q) \leq E[l_k]_p < H(p) + D(p||q) + 1$$

- $\rightarrow$ Nếu chúng ta sử dụng phân bố sai trong quá trình thiết kế mã, thì chúng ta phải trả giá $D(p||q)$ trong độ dài từ mã trung bình mô tả nguồn.

Notes

C3: Mã hóa nguồn - Nén dữ liệu

Nội dung chính

Notes

C3: Mã hóa nguồn - Nén dữ liệu

Nội dung chính

Notes

Một số phương pháp mã hóa nguồn phổ biến

Mã Shannon

Nguyên tắc chọn độ dài từ mã

Với một tin x_k có $p(x_k)$ cho trước, mã Shannon có độ dài từ mã xác định bởi công thức:

$$l_k = \lceil \log_2 \frac{1}{p(x_k)} \rceil \quad (\forall x_k \in X)$$

Thuật toán

1. Sắp xếp các tin theo thứ tự xác suất phân bố giảm dần.
2. Chọn các từ mã có độ dài thích hợp theo thứ tự và tránh việc chọn các từ mã vi phạm tính prefix.

Notes

C3: Mã hóa nguồn - Nén dữ liệu

Nội dung chính

Notes

Một số phương pháp mã hóa nguồn phổ biến

Mã Shannon-Fano

- Thuật toán đơn giản xây dựng bộ mã có tính prefix.
- Thuật toán tạo bộ mã không đều khá hiệu quả (tính toán đơn giản).
- Thuộc lớp thuật toán cận tối ưu (suboptimal).
  - Không luôn luôn tạo ra bộ mã tối ưu.
- Ít phổ biến.

Thuật toán Shannon-Fano

1. Sắp xếp các tin theo thứ tự xác suất (tần suất) từ cao đến thấp từ phía trái sang phía phải.
2. Chia dãy đó thành hai phần sao cho các phần có tổng xác suất xấp xỉ bằng nhau.
3. Gán nhãn cho phần nửa trái một bít 0, và nhóm bên phải bít 1.
4. Lặp lại các bước 3 và 4 cho mỗi nửa bằng cách chia nhóm nhỏ và gán nhãn bít cho đến tận khi các nhóm chỉ còn một nút tương ứng với lá của cây mã.
5. Từ mã thu được bằng cách duyệt từ gốc đến các nút lá tương ứng.

Notes

C3: Mã hóa nguồn - Nén dữ liệu

Nội dung chính

Notes

Một số phương pháp mã hóa nguồn phổ biến

Mã Huffman: Tổng quan

- Thuộc lớp mã hóa Entropy, mã hóa nén dữ liệu không tổn hao (lossless data compression)
- Là lớp mã với độ dài từ mã thay đổi (variable-length code)
- Bộ mã thu được là bộ mã có tính prefix.
- Yêu cầu phân bố của nguồn phải biết trước.
- Thuộc dạng thuật toán "Greedy".
- Là thuật toán mã hóa tối ưu.

Định lý

Mã hóa Huffman là mã hóa tối ưu. Nói cách khác, gọi $\bar{l}_H$ là độ dài trung bình từ mã của bộ mã Huffman cho nguồn rời rạc X, $\bar{l}$ là độ dài trung bình từ mã của bộ mã tạo được bởi một phương pháp nào đó, khi đó chúng ta có:

$$\bar{l}_H \leq \bar{l}$$

Notes

Một số phương pháp mã hóa nguồn phổ biến

Mã Huffman: Bài toán mã hóa

Nhập vào: $X = \{x_k\}$ với các xác suất phân bố $p(x_k)$ tương ứng.

$$X = \{x_k\} = \begin{pmatrix} x_1 & x_2 & \dots & x_N \\ p(x_1) & p(x_2) & \dots & p(x_n) \end{pmatrix}$$

In ra: Các từ mã nhị phân $m_k^{l_k}$ tương ứng với tin x_k

Notes

Một số phương pháp mã hóa nguồn phổ biến

Mã Huffman: Thuật toán mã hóa

1. Khởi động danh sách cây nhị phân có một nút chứa các trọng số là xác suất phân bố tương ứng của các tin x_k, sắp xếp theo thứ tự tăng dần từ trái sang phải.
2. Thực hiện lặp các bước sau đến khi thu được một nút duy nhất.
   1. Tìm hai cây T' và T'' trong danh sách các nút gốc có trọng lượng tối thiểu p' và p''. Thay thế chúng bằng một cây có nút gốc có trọng bằng $p' + p''$ và các cây con là T' và T''.
   2. Đánh nhãn 0 và 1 trên các nhánh từ gốc mới đến các cây T' và T''.
   3. Sắp xếp các nút theo thứ tự tăng dần của trọng xác suất.
3. Duyệt từ gốc cuối cùng đến nút lá với các bít là các nhãn ta được từ mã tương ứng với các tin.

Notes

Một số phương pháp mã hóa nguồn phổ biến

Mã Huffman: Nhận xét

- Phép mã hóa tối ưu Huffman: tập các từ mã cho bộ mã tối ưu là không duy nhất. Nói cách khác, có thể có nhiều hơn một tập các độ dài cho cùng độ dài trung bình:
  - Việc gán nhãn "0" và "1" là tùy ý.
  - Việc sắp xếp các phân bố xác suất hợp (cây thay thế) có thể thực hiện: xếp "trội" nhất, hoặc xếp "chìm" nhất
- Việc xếp xác suất phân bố "trội" nhất sẽ cho bộ mã có phương sai độ dài từ mã nhỏ nhất (gần bộ mã đều nhất)

Mã Huffman thỏa mãn mã tối ưu:

1. Nếu $p(x_k) > p(x_l)$ thì $l_k < l_l$.
2. Hai từ mã có độ dài nhất có cùng độ dài.
3. Hai từ mã có độ dài nhất chỉ khác nhau một bít ở vị trí cuối cùng, và hai từ mã này tương ứng với hai tin (ký hiệu) có xác suất xuất hiện thấp nhất.

Mã Huffman cũng thỏa mãn giới hạn $\bar{l}_k \leq H(X) + 1$

Notes

Một số phương pháp mã hóa nguồn phổ biến

Mã Huffman: Bài toán giải mã, Thuật toán giải mã

Nhập vào: Chuỗi bít thông tin

In ra: Dãy tin tương ứng

1. Khởi động, đặt con trỏ P chỉ đến gốc (root) của cây mã hóa Huffman. Gán con trỏ bít b rỗng.
2. Lặp các bước sau đến khi kết thúc chuỗi bít thông tin
   1. Gán b bằng bít tiếp theo của chuỗi. Nếu $b = 0$ dịch con trỏ P theo nhánh có nhãn 0, nếu ngược lại, dịch con trỏ P theo nhánh có nhãn 1.
   2. Nếu P đã chỉ đến nút lá thì ghi ra tin tương ứng với từ mã. Khởi động lại con trỏ chỉ đến gốc

Notes

C3: Mã hóa nguồn - Nén dữ liệu

Nội dung chính

Notes

Một số phương pháp mã hóa nguồn phổ biến

Mã hóa Shannon-Fano-Elias

1. Sử dụng hàm mật độ phân bố tích lũy để thực hiện mã hóa.
2. Định nghĩa hàm mật độ phân bố tích lũy cải tiến:

$$\bar{F}(x) = \sum_{a<x_k} p(a) + \frac{1}{2}p(x_k)$$

   - $\bar{F}(a) \neq \bar{F}(b)$ nếu $a \neq b$.
   - $\rightarrow$ có thể sử dụng $\bar{F}(x)$ như là một mã cho x_k.

3. Cắt $\bar{F}(x)$ còn l_k bít, ký hiệu là $\lfloor \bar{F} \rfloor_{l_k}$.
4. Nếu $l_k = \lceil \log_2 \frac{1}{p(x_k)} \rceil + 1$ thì:

$$\frac{1}{2^{l_k}} < \frac{p(x_k)}{2} = \bar{F}(x) - F(x-1)$$

   - $\rightarrow$ l_k bít là đủ để có thể mô tả x_k

Notes

C3: Mã hóa nguồn - Nén dữ liệu

Nội dung chính

Notes

Một số phương pháp mã hóa nguồn phổ biến

Mã hóa số học

- Thuộc lớp mã hóa không đều.
- Thuộc lớp mã hóa Entropy.
- Thuộc lớp mã hóa không tổn hao.
- Được sử dụng rộng rãi trong thực tế và trong các trình tiện ích nén dữ liệu thương mại.
- Thực hiện việc mã hóa một nhóm dữ liệu.
- Là một mở rộng trực tiếp của phương pháp mã hóa Shannon-Fano-Elias.
- Ý tưởng quan trọng là tính toán và sử dụng hàm phân bố xác suất của X^n

Notes

C3: Mã hóa nguồn - Nén dữ liệu

Nội dung chính

Notes

Một số phương pháp mã hóa nguồn phổ biến

Thuật toán mã Lempel-Ziv-Welch: Tổng quan

- Thuộc lớp mã hóa không tổn hao.
- Thuộc lớp mã hóa thuật toán từ điển.
- Không yêu cầu phải biết trước phân bố của nguồn, thuật toán thích nghi.
- Ứng dụng rộng rãi trong thực tế, là cơ sở của nhiều trình tiện ích nén dữ liệu thương mại.

Mã hóa Huffman	Mã hóa LZ
Yêu cầu biết phân bố của nguồn	Không cần biết phân bố của nguồn
Bảng mã được chọn trước	Bảng mã được tạo trong quá trình
Phương thức mã độ dài cố định-thay đổi	Phương thức độ dài thay đổi-cố định

Bảng: So sánh giữa mã hóa Huffman và mã hóa LZ. © GIT

Notes

Một số phương pháp mã hóa nguồn phổ biến

Thuật toán mã Lempel-Ziv-Welch: Thuật toán mã hóa

Thuật toán mã hóa Lempel-Ziv

1. Cho trước chuỗi $\mathcal{X} = x_1 x_2 \dots x_n$ (n rất lớn).
2. Khởi động bảng từ mã cơ bản khởi đầu.
3. Tìm kiếm trong chuỗi nguồn đã cho cụm mào đầu dài nhất có mặt trong bảng từ mã. Nói cách khác, tìm kiếm w dài nhất mà $\mathcal{X} = (w, \mathcal{X}')$.
4. Cập nhật bảng mã với từ mã mới được tạo thành từ (w, x_k), với x_k là ký hiệu tiếp theo trong chuỗi đầu vào.

Notes

C3: Mã hóa nguồn - Nén dữ liệu

Nội dung chính

Notes

Kết thúc phần mã hóa nguồn

Notes

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 1: Mã khối tuyến tính)

Lý thuyết thông tin

Biên soạn: Phạm Văn Sự

Bộ môn Xử lý tín hiệu và Truyền thông
Khoa Kỹ thuật Điện tử I
Học viện Công nghệ Bưu chính Viễn thông

ver. 22a

Notes

Mục tiêu của bài học

- Trang bị một số khái niệm cơ bản về mã hóa kênh
- Mã khối tuyến tính
- Mã vòng tuyến tính

Notes

Các câu hỏi cần trả lời

- Các tham số đánh giá mã hóa kênh?
- Khoảng cách mã Hamming tối thiểu? Có vai trò gì trong việc đánh giá khả năng phát hiện lỗi và sửa lỗi của bộ mã?
- Mã khối tuyến tính? Ma trận sinh và ma trận kiểm tra của mã khối tuyến tính? Mã khối tuyến tính hệ thống?
- Bài toán thiết kế mã khối tuyến tính?
- Mã vòng (mã cyclic, mã xyclic) tuyến tính? Đa thức sinh và đa thức kiểm tra của mã vòng tuyến tính? Mã vòng tuyến tính hệ thống?

Notes

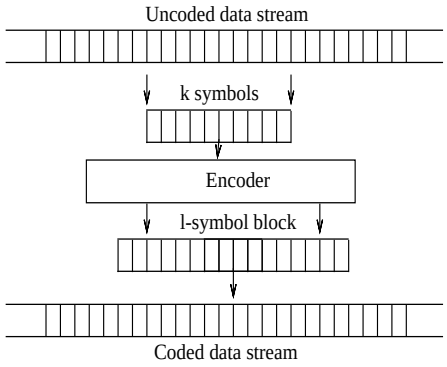
C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 1)

Nội dung chính

Notes

Một số định nghĩa và khái niệm cơ bản

Mã hóa khối



Hình: Quá trình mã hóa khối

Notes

Một số định nghĩa và khái niệm cơ bản

Véc-tơ mã

Định nghĩa (Véc-tơ mã)

Một bộ mã $\mathfrak{C} = \{c_0, c_1, \ldots, c_{M-1}\}$ *chứa các từ mã có độ dài* l*, mỗi từ mã* $c_k = (c_{k,0}, c_{k,1}, \ldots, c_{k,l-1})$ *với các dấu mã* $c_{k,i} \in GF(q)$ $(i = \overline{0, l-1})$.

- $\mathfrak{C}$*: bộ mã cơ số* q
- c_k *được gọi là từ mã, véc-tơ mã*
- M *là số từ mã của bộ mã* $\mathfrak{C}$.

Khối thông tin đầu vào là tập $\{m_i\}$, trong đó $m_i = (m_{i,0}, m_{i,1}, \ldots, m_{i,k-1})$ với $m_{i,j} \in GF(q)$. Tập $\{m_i\}$ tạo thành một không gian véc-tơ trên $GF(q)$.

- Nếu các khối thông tin có cùng độ dài k thì số từ mã của bộ mã $\mathfrak{C}$ phải thỏa mãn $M = q^k$.
- Nếu các khối tin có độ dài thay đổi thì M không có dạng trên.
  - Các bộ mã hóa loại này khó thực thi hơn.

Notes

Một số định nghĩa và khái niệm cơ bản

Độ dư thừa mã, Tỷ số mã, Trọng số mã

Định nghĩa (Độ dư thừa của bộ mã)

Độ dư thừa của bộ mã $\mathfrak{C}$ được định nghĩa là $r = l - \log_q(M)$.

- *Nếu* $M = 2^k$ *thì* $r = l - k$.

Định nghĩa (Tỷ số mã hóa)

Tỷ số mã hóa R được định nghĩa: $R = \frac{\log_q(M)}{l}$

- *Nếu* $M = 2^k$ *thì* $R = k/l$

Định nghĩa (Trọng số của từ mã/cấu trúc lỗi)

Trọng số của một từ mã c *hoặc của một cấu trúc lỗi* e *là số dấu mã khác* 0 *trong* c *hoặc* e. *Kí hiệu là* $w(\mathrm{c})$ *hoặc* $w(\mathrm{e})$

- $0 \leq w(\mathrm{c}) \leq l$

Notes

Một số định nghĩa và khái niệm cơ bản

Khoảng cách mã Hamming

Định nghĩa (Khoảng cách mã Hamming)

Khoảng cách Hamming giữa hai từ mã c_1 *và* c_2 *là tổng số vị trí tương ứng trong hai từ mã mà dấu mã khác nhau.*

$$d_{Hamming}(\mathrm{c}_1, \mathrm{c}_2) = d(\mathrm{c}_1, \mathrm{c}_2) = |\{i | c_{1,i} \neq c_{2,i}, i = 0, 1, ..., l-1\}|$$

- $d(\mathrm{c}_1, \mathrm{c}_2) = d(\mathrm{c}_2, \mathrm{c}_1)$.
- $0 \leq d(\mathrm{c}_1, \mathrm{c}_2) \leq l$.
- $d(\mathrm{c}_1, \mathrm{c}_2) + d(\mathrm{c}_2, \mathrm{c}_3) \geq d(\mathrm{c}_1, \mathrm{c}_3)$ (Bất đẳng thức tam giác).

Định nghĩa (Khoảng cách Hamming tối thiểu)

Khoảng cách mã tối thiểu, hay khoảng cách Hamming tối thiểu của một bộ mã khối $\mathfrak{C}$ *là khoảng cách Hamming tối thiểu giữa tất cả các cặp từ mã phân biệt trong bộ mã.*

$$d_{min} = d_0 = \min_{\forall \mathrm{c}_1, \mathrm{c}_2 \in \mathfrak{C}, \mathrm{c}_1 \neq \mathrm{c}_2} d(\mathrm{c}_1, \mathrm{c}_2)$$

Notes

Một số định nghĩa và khái niệm cơ bản

Khả năng phát hiện và sửa lỗi của mã

Định lý (Khả năng phát hiện lỗi của bộ mã)

Một bộ mã có khoảng cách mã tối thiểu d_{min} có khả năng phát hiện tất cả các cấu trúc lỗi có trọng nhỏ hơn hoặc bằng $(d_{min} - 1)$.

- **Chú ý**: Một số bộ mã có thể phát hiện được các cấu trúc lỗi có trọng $\geq d_{min}$

Định lý (Khả năng sửa lỗi của bộ mã)

Một bộ mã có khoảng cách mã tối thiểu d_{min} có khả năng sửa được tất cả các cấu trúc lỗi có trọng nhỏ hơn hoặc bằng $\lfloor \frac{d_{min}-1}{2} \rfloor$ [a].

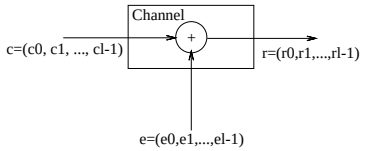
[a] $\lfloor x \rfloor$ là phần nguyên lớn nhất nhỏ hơn hoặc bằng x

- **Chú ý**: Một số bộ mã có thể sửa được các cấu trúc lỗi có trọng $\lfloor \frac{d_{min}-1}{2} \rfloor + 1$ hoặc lớn hơn.

Notes

Một số định nghĩa và khái niệm cơ bản

Mô hình mã truyền dẫn trong kênh có nhiễu



Hình: Mô hình kênh nhiễu cộng

- c: từ mã phát, e: cấu trúc lỗi, r = c + e: véc-tơ thu.
  - Nếu không có lỗi thì véc-tơ thu là một từ mã hợp lệ.
- Định dạng điều chế, mức công suất phát, và mức nhiễu trên kênh quyết định xảy ra một cấu trúc lỗi trong q^l cấu trúc lỗi có thể.
- Máy thu thực hiện việc xem xét véc-tơ thu có phải là từ mã hợp lệ hay không: quá trình phát hiện lỗi.
- Khi máy thu phát hiện lỗi:
  1. Yêu cầu phát lại: thông qua ARQ
  2. HOẶC Đánh dấu từ mã lỗi: với các ứng dụng real-time (voice, video,...)
  3. HOẶC Sửa lỗi: FEC.

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 1)

Nội dung chính

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 1)

Nội dung chính

Notes

Mã khối tuyến tính

Định nghĩa

Định nghĩa (Mã khối tuyến tính)

Xét một bộ mã khối $\mathfrak{C}$ gồm các từ mã độ dài l $\{c_k = (c_{k,0}, c_{k,1}, \ldots, c_{k,l-1})\}$ với các dấu mã thuộc $GF(q)$. Bộ mã $\mathfrak{C}$ là một bộ mã khối tuyến tính cơ số q nếu và chỉ nếu $\mathfrak{C}$ tạo thành một không gian véc-tơ con trên $GF(q)$.

Định nghĩa (Chiều của một bộ mã khối)

Chiều của một bộ mã khối là chiều của không gian véc-tơ tương ứng.

- Ký hiệu: $\mathfrak{C}(l, k)$ hoặc $\mathfrak{C}(l, k, d_0)$.

1. Tổ hợp tuyến tính của một tập các từ mã bất kỳ là một từ mã $\Rightarrow \mathfrak{C}$ luôn chứa từ mã toàn 0
2. Khoảng cách mã tối thiểu của bộ mã khối tuyến tính bằng trọng số của một từ mã có trọng số nhỏ nhất khác từ mã toàn không.
3. Các cấu trúc lỗi không thể phát hiện được của bộ mã độc lập với từ mã phát và luôn chứa tập tất cả các từ mã không toàn 0.

Notes

Mã khối tuyến tính

Ma trận sinh của mã khối tuyến tính

Gọi $\{g_0, g_1, \ldots, g_{k-1}\}$ là cơ sở của các từ mã trong bộ mã $\mathfrak{C}(l, k)$.

Ma trận sinh G$(k \times l)$của bộ mã được thành lập như sau:

$$G = \begin{pmatrix} g_0 \\ g_1 \\ \cdots \\ g_{k-1} \end{pmatrix} = \begin{pmatrix} g_{0,0} & g_{0,1} & \cdots & g_{0,l-1} \\ g_{1,0} & g_{1,1} & \cdots & g_{1,l-1} \\ \vdots & \vdots & \ddots & \vdots \\ g_{k-1,0} & g_{k-1,1} & \cdots & g_{k-1,l-1} \end{pmatrix}$$

Gọi $a = (a_0, a_1, \ldots, a_{k-1})$ là khối dữ liệu đầu vào (bản tin) cần mã hóa.

Từ mã thu được từ phép mã hóa:

$$\begin{aligned} c &= aG = [a_0, a_1, \ldots, a_{k-1}]G \\ &= a_0 g_0 + a_1 g_1 + \cdots + a_{k-1} g_{k-1} \end{aligned}$$

Notes

Mã khối tuyến tính

Ma trận kiểm tra tính chẵn lẻ

Với $\mathfrak{C}$, tồn tại $\mathfrak{C}^{\perp}$ là không gian véc-tơ đối ngẫu $(l-k)$ chiều.

Gọi $\{h_0, h_1, \ldots, h_{l-k-1}\}$ là cơ sở của $\mathfrak{C}^{\perp}$. $\Rightarrow$ Ma trận sinh $H(l-k \times l)$ của $\mathfrak{C}^{\perp}$:

$$H = \begin{pmatrix} h_0 \\ h_1 \\ \ldots \\ h_{l-k-1} \end{pmatrix} = \begin{pmatrix} h_{0,0} & h_{0,1} & \ldots & h_{0,l-1} \\ h_{1,0} & h_{1,1} & \ldots & h_{1,l-1} \\ \vdots & \vdots & \ddots & \vdots \\ h_{l-k-1,0} & h_{l-k-1,1} & \ldots & h_{l-k-1,l-1} \end{pmatrix}$$

- H là ma trận kiểm tra chẵn lẻ của mã $\mathfrak{C}$
- $GH^T = 0$.

Định lý

Một véc-tơ c *là một từ mã thuộc* $\mathfrak{C}$ *nếu và chỉ nếu* $cH^T = 0$

- $cH^T = 0$ gọi là biểu thức kiểm tra chẵn lẻ.

Notes

Mã khối tuyến tính

Ma trận kiểm tra tính chẵn lẻ và khoảng cách mã

Định lý

Giả sử bộ mã $\mathfrak{C}$ *có ma trận kiểm tra tính chẵn lẻ* H. *Khoảng cách mã tối thiểu của bộ mã* $\mathfrak{C}$ *bằng số cột tối thiểu khác* 0 *của* H *mà tổ hợp tuyến tính không tầm thường của chúng bằng* 0.

Định lý (Giới hạn Singleton)

Với bộ mã khối tuyến tính $\mathfrak{C}(l,k)$, *khoảng cách mã tối thiểu thỏa mãn bất đẳng thức:*

$$d_{min} \leq l - k + 1$$

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 1)

Nội dung chính

Notes

Mã khối tuyến tính

Mã khối tuyến tính hệ thống

Định nghĩa (Mã khối tuyến tính hệ thống)

Mã khối tuyến tính hệ thống $\mathfrak{C}(l,k)$ thực hiện việc ánh xạ bản tin (khối dữ liệu) độ dài k thành một véc-tơ/từ mã độ dài l sao cho trong số l bít có thể chỉ ra k bít bản tin và số còn lại $l-k$ bít kiểm tra tính chẵn lẻ.

Giả sử từ mã xây dựng mã có dạng $c = [p_1 \mid a]$

- a: khối thông tin (bản tin) độ dài k; p_1: khối bít kiểm tra độ dài $l-k$

$G \xrightarrow{\text{phương pháp khử Gausse}} G = [P \mid I_k]$

- $P_{(k\times l-k)}$: ma trận tạo dấu kiểm tra
- $I_{(k\times k)}$: ma trận đơn vị.

$\Rightarrow H = [I_{l-k} \mid -P^T]$

- $\Rightarrow GH^T = 0$

Chý ý: Nếu xét $c = [a \mid p_1]$

- $G = [I_k \mid P]$
- $\Rightarrow H = [-P^T \mid I_{l-k}]$

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 1)

Nội dung chính

Notes

Đánh giá mã khối nhị phân tuyến tính trên kênh BSC

Ví dụ

Ví dụ

Xét bộ mã nhị phân đều chiều dài l (ví dụ bộ mã nhị phân đều chiều dài 2: $\mathfrak{C} = (00), (01), (11), (10)$). Giả sử kết quả mã hóa được truyền qua kênh nhị phân rời rạc đối xứng không nhớ (BSC) có xác suất thu sai p_0, các bít được phát đi độc lập nhau, và xác suất phát đi bít 0 và bít 1 tương đương nhau.

1. *Tính xác suất thu được một từ mã đúng.*
2. *Giả sử xác suất sai cho phép đối với việc thu các từ mã là p_a, tìm điều kiện đối với p_0 để có thể sử dụng được bộ mã cho việc thông tin qua kênh.*

Notes

Đánh giá mã khối nhị phân tuyến tính trên kênh BSC

Đánh giá khả năng phát hiện lỗi

Cho $\mathfrak{C}(l, k, d_{min})$ truyền qua kênh BSC có xác suất chuyển sai p.

- $P_u(E)$: xác suất véc-tơ thu có lỗi mà không phát hiện được.
- $P_e(E)$: xác suất véc-tơ thu có lỗi.
- $P_d(E)$: xác suất véc-tơ thu có lỗi được phát hiện.

$$
\begin{aligned}
P_u(E) &\leq \sum_{j=d_{min}}^{l} \binom{l}{j} p^j (1-p)^{l-j} = 1 - \sum_{j=0}^{d_{min}-1} \binom{l}{j} p^j (1-p)^{l-j} \\
P_u(E) &= \sum_{j=d_{min}}^{l} A_j p^j (1-p)^{l-j} \\
P_e(E) &= \sum_{j=1}^{l} \binom{l}{j} p^j (1-p)^{l-j} = 1 - (1-p)^l \\
P_d(E) &= P_e(E) - P_u(E) = 1 - (1-p)^l - P_u(E)
\end{aligned}
$$

Notes

Đánh giá mã khối nhị phân tuyến tính trên kênh BSC

Đánh giá khả năng phát hiện lỗi (cont.)

- P_{u_b}: tỷ lệ bít lỗi không được phát hiện
  - $\triangleq$ xác suất bít thông tin nhận được bị lỗi trong một từ mã bị tác động bởi cấu trúc lỗi không phát hiện được
  - $P_u(E) \geq P_{u_b}(E) \geq \frac{1}{k} P_u(E)$
- P_{d_b}: tỷ lệ bít lỗi được phát hiện
  - $\triangleq$ xác suất bít thông tin nhận được bị lỗi trong một từ mã bị tác động bởi cấu trúc lỗi có thể phát hiện được.
  - $P_d(E) \geq P_{d_b}(E) \geq \frac{1}{k} P_d(E)$
- Nếu biết phân bố trọng của bộ mã, P_{u_b} có thể tính một cách chính xác:

$$P_{u_b} = \sum_{j=d_{min}}^{l} \frac{B_j}{k} p^j (1-p)^{l-j}$$

trong đó B_j là tổng trọng của các khối tin tương ứng với tất cả các từ mã có trọng là i.

Notes

Đánh giá mã khối nhị phân tuyến tính trên kênh BSC

Đánh giá khả năng sửa lỗi

Cho $\mathfrak{C}(l, k, d_{min})$ truyền qua kênh BSC có xác suất chuyển sai p.

Xét bộ giải mã có độ dài giới hạn.

- $P(E)$: xác suất giải mã sai

$$P(E) \leq \sum_{j=\lfloor \frac{d_{min}-1}{2} \rfloor + 1}^{l} \binom{l}{j} p^j (1-p)^{l-j} = 1 - \sum_{j=0}^{\lfloor \frac{d_{min}-1}{2} \rfloor} \binom{l}{j} p^j (1-p)^{l-j}$$

Đẳng thức xảy ra chỉ khi mã là hoàn hảo.

- $P(F)$: xác suất giải mã thất bại

$$P(F) \leq 1 - \sum_{j=0}^{\lfloor \frac{d_{min}-1}{2} \rfloor} \binom{l}{j} p^j (1-p)^{l-j}$$

Notes

Đánh giá mã khối nhị phân tuyến tính trên kênh BSC

Đánh giá khả năng sửa lỗi (cont')

Xét $\mathfrak{C}(l, k, d_{min})$ với phân bố trọng số đã biết $\{A_i\}$

$P_k^j \triangleq$ xác suất một véc-tơ thu có khoảng cách Hamming chính xác là k so với một từ mã có trọng là j.

$$P_k^j = \sum_{r=0}^{k} \binom{j}{k-r} \binom{l-j}{r} p^{j-k+2r} (1-p)^{l-j+k-2r}$$

$$\begin{aligned} P(E) &= \sum_{j=d_{min}}^{l} A_j \sum_{k=0}^{\lfloor \frac{d_{min}-1}{2} \rfloor} P_k^j \\ P(F) &= 1 - \sum_{j=0}^{\lfloor \frac{d_{min}-1}{2} \rfloor} \binom{l}{j} p^j (1-p)^{l-j} - P(E) \end{aligned}$$

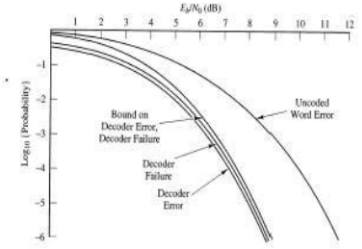
Notes

Đánh giá mã khối nhị phân tuyến tính trên kênh BSC

Đánh giá khả năng sửa lỗi (cont')

- Nếu biết được mối quan hệ giữa trọng số của các khối tin và trọng số các từ mã tương ứng
  - $\Rightarrow B_j$
- $\Rightarrow$

$$BER = P_b(E) = \frac{1}{k}\sum_{j=d_{min}}^{l} B_j \sum_{k=0}^{\lfloor \frac{d_{min}-1}{2} \rfloor} P_k^j$$



Chú ý: Thường, thông tin $\{B_j\}$ không khả thi.

- $\Rightarrow$ Chủ yếu dựa vào các đánh giá biên

$$P(E) \geq P_b(E) \geq \frac{1}{k}P(E)$$

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 1)

Nội dung chính

Notes

Các vấn đề khi thiết kế mã khối tuyến tính

Thiết kế mã khối tuyến tính tối ưu

Khi thiết kế, ta mong muốn có được bộ mã có độ dư thừa nhỏ nhất có thể, nhưng lại có khả năng phát hiện và sửa lỗi lớn nhất có thể.

Trường hợp 1

Với k và d_{min} cho trước, xây dựng bộ mã có độ dư thừa tối thiểu: $\min\{l\}$.
Độ dài từ mã của bộ mã thỏa mãn giới hạn Griesmer:

$$l \geq \sum_{i=0}^{k-1} \lceil \frac{d_{min}}{2^i} \rceil$$

$\lceil x \rceil$: phần nguyên nhỏ nhất lớn hơn hoặc bằng x.

Notes

Các vấn đề khi thiết kế mã khối tuyến tính

Thiết kế mã khối tuyến tính tối ưu (cont')

Trường hợp 2

Với l và k cho trước, xây dựng bộ mã có khả năng phát hiện và sửa sai lớn nhất: $\max\{d_{min}\}$.
Khoảng cách Hamming tối thiểu của bộ mã thỏa mãn giới hạn Plotkin:

$$d_{min} \leq \frac{l \times 2^{k-1}}{2^k - 1}$$

Trường hợp 3

Với l và khả năng sửa sai t cho trước, xây dựng bộ mã có độ dư thừa nhỏ nhất: $\max\{k\}$.
Mối liên hệ giữa l, k và t thỏa mãn giới hạn Hamming:

$$2^{l-k} \geq \sum_{i=0}^{t} \binom{l}{i}$$

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 1)

Nội dung chính

Notes

Kết thúc phần mã khối tuyến tính

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2: Mã vòng tuyến tính)

Lý thuyết thông tin

Biên soạn: Phạm Văn Sự

Bộ môn Xử lý tín hiệu và Truyền thông
Khoa Kỹ thuật Điện tử I
Học viện Công nghệ Bưu chính Viễn thông

ver 22a

Notes

Mục tiêu của bài học

- Tiếp tục trang bị một số khái niệm cơ bản về mã hóa kênh
- Mã vòng (mã cyclic, mã xyclic) tuyến tính

Notes

Các câu hỏi cần trả lời

- Vành đa thức đồng dư?
- Đa thức sinh, đa thức kiểm tra của mã vòng tuyến tính?
- Mã vòng tuyến tính hệ thống? Thuật toán lập mã cho mã vòng tuyến tính hệ thống?
- Các phương pháp giải mã cơ bản cho mã vòng tuyến tính?

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2)

Nội dung chính

Notes

Đa thức mã và các phép biến đổi

Đa thức mã

Véc-tơ mã $c = (c_0, c_1, \ldots, c_{l-1})$ có thể biểu diễn ở dạng đa thức:

$$c(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{l-1} x^{l-1}$$

Nhận xét:

- Mỗi véc-tơ mã/từ mã có chiều dài l tương ứng với một đa thức bậc nhỏ hơn hoặc bằng $l-1$.
- Mối quan hệ giữa véc-tơ mã với biểu diễn đa thức đảm bảo $1-1$.
- $c(x)$ gọi là đa thức mã. Khái niệm từ mã/véc-tơ mã và đa thức mã có thể được dùng thay thế nhau.
  - $c \in \mathfrak{C}(l,k) \Leftrightarrow c(x) \in GF(q)[x]/(x^l - 1)$

Notes

Đa thức mã và các phép biến đổi

Phép cộng đa thức, Phép nhân đa thức

- Xét các đa thức $f(x)$, $g(x)$ trên $GF(q)[x]/(x^l - 1)$

Phép cộng đa thức

$f(x) = f_0 + f_1 x + f_2 x^2 + \cdots + f_{l-1} x^{l-1}$
$g(x) = g_0 + g_1 x + g_2 x^2 + \cdots + g_{l-1} x^{l-1}$
$\Rightarrow f(x) + g(x) = (f_0 + g_0) + (f_1 + g_1)x + \cdots + (f_{l-1} + g_{l-1})x^{l-1}$

Phép nhân đa thức

$f(x) = f_0 + f_1 x + f_2 x^2 + \cdots + f_{l-1} x^{l-1} = \sum_{i=0}^{l-1} f_i x^i$
$g(x) = g_0 + g_1 x + g_2 x^2 + \cdots + g_{l-1} x^{l-1} = \sum_{j=0}^{l-1} g_j x^j$
$\Rightarrow f(x) \times g(x) = (\sum_{i=0}^{l-1} f_i x^i)(\sum_{j=0}^{l-1} g_j x^j)$ modulo $(x^l - 1)$

Notes

Đa thức mã và các phép biến đổi

Phép dịch vòng

Trên $GF(q)[x]/(x^l - 1)$, cho $f(x) = \sum_{i=0}^{l-1} f_i x^i \longleftrightarrow \mathrm{a} = (f_0, f_1, \ldots, f_{l-1})$

Xét $g(x) = x.f(x) \longleftrightarrow \mathrm{b} = (f_{l-1}, f_0, f_1, \ldots, f_{l-2})$ (chú ý: mod $x^l - 1$)

- b thu được bằng cách dịch vòng về phía phải của a một cấp/nhịp/vòng.
- Kí hiệu $g(x) = f^{(1)}(x)$.
- $\Rightarrow$ Nhân x^i với $f(x)$ thu được một véc-tơ là kết quả dịch vòng phải của véc-tơ ban đầu đi i nhịp/cấp: $f^{(i)}(x)$.

Xét $g(x) = \frac{f(x)}{x} \longleftrightarrow \mathrm{b} = (f_1, f_2, f_3, \ldots, f_{l-1}, f_0)$ (chú ý: mod $x^l - 1$)

- b thu được bằng cách dịch vòng về phía trái của a một cấp/vòng.
- $\Rightarrow$ Chia $f(x)$ cho x^i thu được một véc-tơ là kết quả dịch vòng trái của véc-tơ ban đầu đi i nhịp/cấp.

Notes

Đa thức mã và các phép biến đổi

Đa thức đối ngẫu

Định nghĩa

Cho đa thức $f(x)$ bậc k: $f(x) = f_0 + f_1 x + f_2 x^2 + \cdots + f_k x^k$.
Đa thức đối ngẫu của $f(x)$, kí hiệu là $f^*(x)$ được định nghĩa là:

$$f^*(x) = x^k \times f(x^{-1}) = f_k + f_{k-1} x + f_{k-2} x^2 + \cdots + f_1 x^{k-1} + f_0 x^k$$

- Nếu $f^*(x) = f(x)$ thì $f(x)$ là đa thức tự đối ngẫu.

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2)

Nội dung chính

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2)

Nội dung chính

1. Đa thức mã và các phép biến đổi
2. Mã vòng tuyến tính
   - Một số định nghĩa và khái niệm
   - Ma trận sinh và ma trận kiểm tra của mã vòng
   - Mã vòng tuyến tính dạng hệ thống
3. Mạch nguyên lý mã hóa mã vòng
   - Xây dựng từ đa thức sinh
   - Xây dựng từ đa thức kiểm tra
4. Các phương pháp giải mã vòng
   - Phương pháp giải mã ngưỡng
   - Phương pháp bẫy lỗi - Thuật toán chia dịch vòng
5. Kết thúc

Notes

Mã vòng tuyến tính

Một số định nghĩa và khái niệm: Định nghĩa

Định nghĩa

Một mã khối tuyến tính $\mathfrak{C}(l,k)$ được gọi là mã vòng tuyến tính nếu với mọi từ mã $c = (c_0, c_1, \ldots, c_{l-1}) \in \mathfrak{C}$ thì kết quả của mỗi dịch vòng từ mã c cũng sẽ thu được một véc-tơ cũng là một từ mã thuộc $\mathfrak{C}$.

Cho $a(x) \in GF(q)[x]/(x^l - 1)$, $c(x) \in \mathfrak{C}$

$\Rightarrow a(x)c(x)$ là tổ hợp tuyến tính của các dịch vòng của $c(x)$

$\Rightarrow a(x)c(x) \in \mathfrak{C}\ \forall a(x) \in GF(q)[x]/(x^l - 1)$, $c(x) \in \mathfrak{C}$

Notes

Mã vòng tuyến tính

Một số định nghĩa và khái niệm: Một số tính chất, Đa thức sinh

Định lý

Bộ mã $\mathfrak{C}$ là một bộ mã vòng tuyến tính cơ số q có chiều dài từ mã l nếu và chỉ nếu các đa thức mã của $\mathfrak{C}$ tạo thành một ideal trên $GF(q)[x]/(x^l - 1)$.

- Trong tập tất cả các đa thức mã của $\mathfrak{C}$, có một đa thức monic duy nhất $g(x)$ với bậc tối thiểu $r = l - k < l$. $g(x)$ được gọi là đa thức sinh của bộ mã $\mathfrak{C}$.
- Mọi đa thức mã $c(x) \in \mathfrak{C}$ tồn tại duy nhất một biểu diễn $c(x) = a(x)g(x)$, trong đó $g(x)$ là đa thức sinh, $a(x)$ là đa thức bậc $\leq l - r = k$ trên $GF(q)[x]$.
- Đa thức sinh $g(x)$ của bộ mã $\mathfrak{C}$ là một thừa số của $x^l - 1$ trên $GF(q)[x]$.

Định lý

Nếu $g(x)$ có bậc $r = l - k$ và là một thừa số của $x^l - 1$ thì $g(x)$ là một đa thức sinh của mã vòng tuyến tính $\mathfrak{C}(l,k)$.

Notes

Mã vòng tuyến tính

Một số định nghĩa và khái niệm: Đa thức kiểm tra, Mã vòng đối ngẫu

Định nghĩa

Một bộ mã vòng tuyến tính $\mathfrak{C}(l,k)$ có đa thức sinh $g(x)$. Một đa thức $h(x) \neq 0$ được gọi là đa thức kiểm tra của $\mathfrak{C}(l,k)$ nếu $g(x) \times h(x) = x^l - 1 \equiv 0$ (mod $x^l - 1$)

- $deg(h(x)) = k$
- $h(x) = \frac{x^l-1}{g(x)}$

Định lý

$\mathfrak{C}(l,k)$ là một mã vòng tuyến tính với đa thức sinh $g(x)$. Khi đó, mã đối ngẫu $\mathfrak{C}^{\perp}$ cũng là một mã vòng tuyến tính $(l, l-k)$ và được sinh ra từ đa thức sinh $h^*(x) = x^k h(x^{-1})$ với $h(x) = \frac{(x^l-1)}{g(x)}$.

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2)

Nội dung chính

Notes

Mã vòng tuyến tính

Một số định nghĩa và khái niệm: Ma trận sinh của mã vòng

Một bộ mã vòng tuyến tính $\mathfrak{C}(l,k)$ với đa thức sinh $g(x) = g_0 + g_1x + g_2x^2 + \cdots + g_{l-k}x^{l-k}$ có ma trận sinh xác định bởi:

$$\mathrm{G} = \begin{bmatrix} g(x) \\ xg(x) \\ x^2g(x) \\ \vdots \\ x^{k-1}g(x) \end{bmatrix} = \begin{bmatrix} g_0 & g_1 & g_2 & \dots & g_{l-k} & 0 & 0 & \dots & 0 \\ 0 & g_0 & g_1 & \dots & g_{l-k-1} & g_{l-k} & 0 & \dots & 0 \\ 0 & 0 & g_0 & \dots & g_{l-k-2} & g_{l-k-1} & g_{l-k} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \dots & \dots & \dots & \dots & g_{l-k} \end{bmatrix}$$

- G có kích thước $k \times l$
- G không có dạng hệ thống

Notes

Mã vòng tuyến tính

Một số định nghĩa và khái niệm: Ma trận kiểm tra của mã vòng

Trên $GF(q)$, xét bộ mã vòng tuyến tính $\mathfrak{C}(l,k)$ với đa thức sinh $g(x)$. Tồn tại một đa thức $h(x)$ bậc $k = l - r$ thỏa mãn $g(x)h(x) = x^l - 1$, hay $h(x)g(x) \equiv 0 \bmod x^l - 1$. $h(x)$ được gọi là đa thức kiểm tra của mã $\mathfrak{C}(l,k)$.

Xét một bộ mã vòng tuyến tính $\mathfrak{C}(l,k)$ với đa thức kiểm tra $h(x) = h_0 + h_1x + h_2x^2 + \cdots + h_kx^k$, ma trận kiểm tra của nó được xác định bởi:

$$\mathrm{H} = \begin{bmatrix} h_k & h_{k-1} & h_{k-2} & \dots & h_0 & 0 & 0 & \dots & 0 \\ 0 & h_k & h_{k-1} & \dots & h_1 & h_0 & 0 & \dots & 0 \\ 0 & 0 & h_k & \dots & h_2 & h_1 & h_0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \dots & \dots & \dots & \dots & h_0 \end{bmatrix}$$

- H có kích thước $l - k \times l$
- $\mathrm{GH}^T = 0$

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2)

Nội dung chính

Notes

Mã vòng tuyến tính dạng hệ thống

Thuật toán chia = Thuật toán bốn bước = Thuật toán tạo từ mã dạng hệ thống từ đa thức sinh

Từ mã dạng hệ thống $c = [p \mid a]$

Bài toán

Nhập vào: $\mathfrak{C}(l, k)$, $g(x)$, khối tin cần mã hóa $a = (a_0, a_1, \ldots, a_{k-1})$.
In ra: Từ mã dạng hệ thống tương ứng c

Thuật toán

1. Mô tả khối tin bằng biểu diễn đa thức tương ứng $a(x)$.
2. Tính $a^{(l-k)}(x) = x^{l-k}a(x)$.
3. Chia $x^{l-k}a(x)$ cho đa thức sinh $g(x)$ của bộ mã, thu được phần dư $p(x)$.
4. Thành lập đa thức mã $c(x) = p(x) + x^{l-k}a(x)$. In ra từ mã tương ứng với đa thức mã $c(x)$.

Notes

Mã vòng tuyến tính dạng hệ thống

Thuật toán nhân = Thuật toán tạo từ mã dạng hệ thống từ đa thức kiểm tra

- Hoàn toàn có thể xây dựng được mã vòng tuyến tính dạng hệ thống từ đa thức (ma trận) kiểm tra.

Xây dựng mã hệ thống từ đa thức kiểm tra

1. Từ khối tin vào (tương ứng đa thức tin) ta có: $c_{l-k} = a_0$, $c_{l-k+1} = a_1$, ..., $c_{l-1} = a_{k-1}$.
2. Tính toán c_0, c_1, ..., c_{l-k-1} từ công thức:

$$c_{l-k-i} = \sum_{j=0}^{k-1} h_j c_{l-j-i} \quad (1 \le i \le l-k)$$

3. Từ mã tương ứng dạng hệ thống $c = (c_0, c_1, c_2, \ldots, c_{l-k-1}, a_0, \ldots, a_{k-1})$.

Notes

Mã vòng tuyến tính dạng hệ thống

Ma trận sinh, ma trận kiểm tra dạng hệ thống

$$G \xrightarrow{\text{phương pháp khử Gausse}} G_{\text{dạng hệ thống}}$$

- Nếu $G = \begin{bmatrix} P & | & I_k \end{bmatrix} \Rightarrow H = \begin{bmatrix} I_{l-k} & | & P^T \end{bmatrix}$
- Nếu $G = \begin{bmatrix} I_k & | & P \end{bmatrix} \Rightarrow H = \begin{bmatrix} P^T & | & I_{l-k} \end{bmatrix}$

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2)

Nội dung chính

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2)

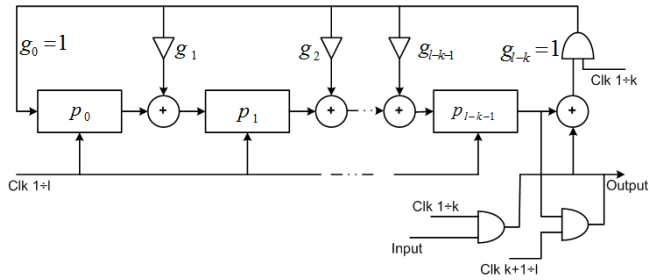
Nội dung chính

1. Da thức mã và các phép biến đổi
2. Mã vòng tuyến tính
   - Một số định nghĩa và khái niệm
   - Ma trận sinh và ma trận kiểm tra của mã vòng
   - Mã vòng tuyến tính dạng hệ thống
3. Mạch nguyên lý mã hóa mã vòng
   - Xây dựng từ đa thức sinh
   - Xây dựng từ đa thức kiểm tra
4. Các phương pháp giải mã vòng
   - Phương pháp giải mã ngưỡng
   - Phương pháp bẫy lỗi - Thuật toán chia dịch vòng
5. Kết thúc

Notes

Mã vòng tuyến tính dạng hệ thống

Mạch nguyên lý mã hóa mã vòng: Xây dựng từ đa thức sinh - Sơ đồ mạch nguyên lý



Hình: Mạch thực hiện mã hóa mã vòng dạng tuyến tính dựa trên đa thức sinh

Notes

Mã vòng tuyến tính dạng hệ thống

Mạch nguyên lý mã hóa mã vòng: Xây dựng từ đa thức sinh - Nguyên lý hoạt động

1. Đầu tiên, nội dung các thanh ghi được xóa về 0.
2. k nhịp đầu tiên, véc-tơ tin (a) được dịch trực tiếp ra đầu ra và đồng thời được dịch vào mạch để tính các bít kiểm tra. Sau k nhịp, nội dung các thanh ghi là các bít kiểm tra.
3. $l - k$ nhịp tiếp theo, mạch thực hiện dịch nội dung các bít kiểm tra trong thanh ghi ra đầu ra.
4. Quá trình mã hóa kết thúc khi toàn bộ khối bít kiểm tra được dịch ra ngoài.

Notes

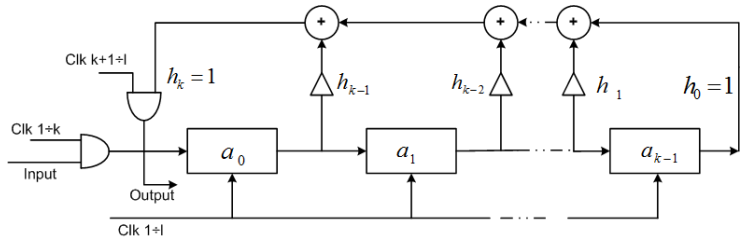
C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2)

Nội dung chính

Notes

Mã vòng tuyến tính dạng hệ thống

Mạch nguyên lý mã hóa mã vòng: Xây dựng từ đa thức kiểm tra - Mạch nguyên lý



Hình: Sơ đồ mạch mã hóa mã vòng dạng hệ thống dựa trên đa thức kiểm tra

Notes

Mã vòng tuyến tính dạng hệ thống

Mạch nguyên lý mã hóa mã vòng: Xây dựng từ đa thức kiểm tra - Nguyên lý hoạt động

1. Đầu tiên, nội dung các thanh ghi thông tin được xóa về 0.
2. k nhịp đầu tiên, khối thông tin được dịch vào các thanh ghi đồng thời dịch ra đầu ra. Sau k nhịp, nội dung các thanh ghi là nội dung của khối tin.
3. $l-k$ nhịp tiếp theo, các c_{l-k-i} $(i=\overline{1,l-k})$ được tính và được chuyển vào thanh ghi đồng thời chuyển ra đầu ra.
4. Quá trình mã hóa kết thúc sau khi $l-k$ bít kiểm tra được lập xong.

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2)

Nội dung chính

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2)

Nội dung chính

Notes

Các phương pháp giải mã vòng

Phương pháp giải mã ngưỡng: Hệ tổng kiểm tra

$c \in \mathfrak{C}$, $w \in \mathfrak{C}^{\perp}$ $A \triangleq wr = we$

A: một tổng kiểm tra.

$A = w_0 e_0 + w_1 e_1 + \cdots + w_{l-1} e_{l-1}$

- bít lỗi e_k được kiểm tra bằng tổng kiểm tra A nếu $w_k = 1$.

Định nghĩa (Hệ tổng kiểm tra trực giao)

Một hệ gồm J tổng kiểm tra được gọi là hệ tổng kiểm tra trực giao với vị trí bít lỗi e_{l-1} nếu:

1. *Tất cả các hệ số của e_{l-1} trong hệ J tổng kiểm tra bằng 1.*
2. *Với $k \neq l-1$ chỉ có nhiều nhất một véc-tơ trong hệ tổng kiểm tra mà hệ số của e_k bằng 1.*

$\Rightarrow A_k = e_{l-1} + \sum_{i \neq l-1} w_i e_i$

Notes

Các phương pháp giải mã vòng

Phương pháp giải mã ngưỡng: Thuật toán một bước

Giải mã ngưỡng dựa trên hệ tổng kiểm tra trực giao

Bít lỗi e_{l-1} được quyết định là 1 nếu có phần lớn các véc-tơ trong tổng kiểm tra trực giao bằng 1. Ngược lại thì bít lỗi e_{l-1} được quyết định là 0.

- Bộ giải mã hoạt động đúng khi véc-tơ lỗi có trọng $\leq \lfloor J/2 \rfloor$.
- Nếu có thể tạo hệ J tổng kiểm tra trực giao cho e_{l-1} thì cũng có thể tạo hệ J tổng kiểm tra trực giao cho các vị trí bít lỗi e_k $(k \neq l-1)$ nào đó.
- Nếu J là số tổng kiểm tra trực giao cực đại có thể lập được cho e_{l-1} (hoặc bất kỳ e_k nào đó), phương pháp giải mã nêu trên có thể sửa được các cấu trúc lỗi có trọng $\leq \lfloor J/2 \rfloor$. $t_{ML} = \lfloor J/2 \rfloor$: khả năng sửa lỗi của bộ giải mã ngưỡng.
- Phép giải mã này được gọi là hiệu quả với bộ mã $\mathfrak{C}(l, k, d_0)$ chỉ nếu $t_{ML} = \lfloor J/2 \rfloor$ bằng hoặc xấp xỉ bằng $t = \lfloor (d_0 - 1)/2 \rfloor$.

Notes

Các phương pháp giải mã vòng

Phương pháp giải mã ngưỡng: Hệ tổng kiểm tra có khả năng trực giao

Định nghĩa (Bộ mã vòng có khả năng trực giao đầy đủ)

Một bộ mã vòng $\mathfrak{C}(l, k, d_0)$ gọi là có khả năng trực giao đầy đủ một bước nếu và chỉ nếu nó có thể tạo được hệ $J = d_0 - 1$ tổng kiểm tra trực giao với một vị trí bít lỗi nào đó.

- $J < l - k$.
- Không phải mọi mã vòng $\mathfrak{C}(l, k, d_0)$ đều là có khả năng trực giao đầy đủ.

Định nghĩa (Hệ tổng kiểm tra có khả năng trực giao)

Một tập gồm J tổng kiểm tra $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_J$ là hệ tổng kiểm tra trực giao với tập M vị trí bít lỗi $\mathrm{E} = \{e_{i_1}, e_{i_2}, \ldots, e_{i_M}\}$ $(0 \leq i_1 < i_2 < \cdots < i_M < l)$ nếu:

1. *Mọi vị trí bít lỗi e_{i_j} của E đều được kiểm tra bởi mọi tổng kiểm tra A_j $(1 \leq j \leq J)$, và*
2. *Không có bất cứ vị trí lỗi nào khác được kiểm tra ở nhiều hơn 1 tổng kiểm tra.*

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2)

Nội dung chính

Notes

Các phương pháp giải mã vòng

Phương pháp bẫy lỗi - Thuật toán chia dịch vòng: Thuật toán

Nhập vào: Véc-tơ thu $r(x)$ và thông số bộ mã $\mathfrak{C}(l,k)$ như đa thức sinh $g(x)$ và d_{min}, kí hiệu $\mathfrak{C}(l,k,d_{min})$.
In ra Từ mã đã được sửa sai.

Bước 1: Với $i = 0,...,l-1$

1. Tính $s_i(x)$ là phần dư của phép chia $x^i r(x)$ [hoặc $\frac{r(x)}{x^i}$] cho $g(x)$.
2. Tính trọng của $s_i(x)$: $w(s_i(x))$.
3. Nếu $w(s_i(x)) \leq t = \lfloor \frac{d_{min}-1}{2} \rfloor$ chuyển đến **Bước 2**.
4. Nếu $w(s_i(x)) > t$ tăng i lên 1 đơn vị.
5. Nếu $i = l$ chuyển đến **Bước 3**.

Bước 2 Đa thức mã được sửa bởi: $\hat{r}(x) = \frac{x^i r(x)+s_i(x)}{x^i}$ [hoặc $\hat{r}(x) = x^i\{\frac{r(x)}{x^i} + s_i(x)\}$]. In ra từ mã đã được sửa lỗi tương ứng. Kết thúc.
Bước 3 Thông báo không sửa được lỗi (số lỗi vượt quá khả năng sửa lỗi). Kết thúc.

Notes

C4: Mã hóa kênh - Truyền dẫn dữ liệu (Part 2)

Nội dung chính

Notes

Kết thúc phần mã vòng

Notes