Teoria da Informação

Charles Casimiro Cavalcante

charles@gtel.ufc.br

Grupo de Pesquisa em Telecomunicações Sem Fio – GTEL
Programa de Pós-Graduação em Engenharia de Teleinformática
Universidade Federal do Ceará – UFC
http://www.gtel.ufc.br/∼charles

"A principal função de um sistema de comunicação é reproduzir, exatamente ou de forma aproximada, uma informação proveniente de outro ponto diferente."

Claude Shannon, 1948

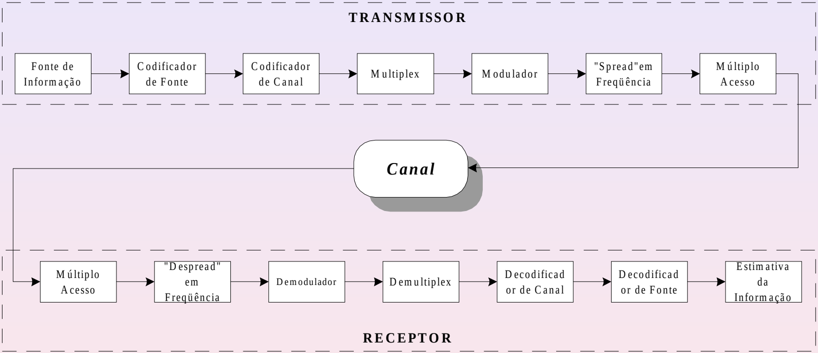
Conteúdo do curso

1. Revisão de probabilidade
2. Informação e Entropia
3. Codificação de fontes
4. Codificação e capacidade de canal
5. Complexidade de Kolmogorov
6. Funções de otimização
7. *Independent Component Analysis*

Parte IV

Codificação e Capacidade de Canal

Sistema de comunicação - modelo geral



Considerações gerais

- Codificador de fonte *retira* redundância - compactação da informação
- Canal insere perturbações/interferência
- Necessidade de **inserção de redundância** (conhecida) para evitar/minimizar os efeitos do canal
- Transferência de informação torna-se complicada
- Necessidade de inclusão de algum **controle de qualidade** na mensagem

Pergunta

Qual o limite de informação que pode ser transmitida, sem erros, através de um canal que insere perturbação na mensagem?

Resposta

- Resposta a pergunta anterior: teorema da capacidade de canal

Teorema da Capacidade de Canal

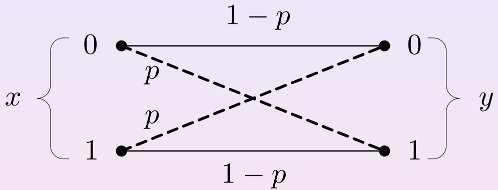
A capacidade de informação de um canal é a quantidade máxima de informação capaz de ser transmitida, *bits por uso do canal*, sem erro de transmissão. Define-se então:

$$C = \max_{p_X(x)} \mathcal{I}(X,Y) \tag{90}$$

- Capacidade de canal é uma função apenas das probabilidades condicionais $P(y_k|x_j)$ que definem o canal
- A maximização em relação às probabilidades dos símbolos (codificados) da fonte é sujeita a
    1. $P(x_i) \geq 0$, para todo i
    2. $\sum_{i=0}^{K-1} P(x_i) = 1$

Capacidade de canal - propriedades

1. $C \geq 0$, uma vez que $\mathcal{I}(X,Y) \geq 0$
2. $C \leq \log[|X|]$, uma vez que
   $C = \max\ \mathcal{I}(X,Y) \leq \max\ \mathcal{H}(X) = \log[|X|]$
3. $C \leq \log[|Y|]$
4. $\mathcal{I}(X,Y)$ é uma função contínua de $p_X(x)$
5. $\mathcal{I}(X,Y)$ é uma função côncava de $p_X(x)$



- Diagrama é unicamente definido pela probabilidade de transição
- A entropia é maximizada quando $\Pr(x_0) = \Pr(x_1) = 1/2$

Capacidade de canal - exemplo

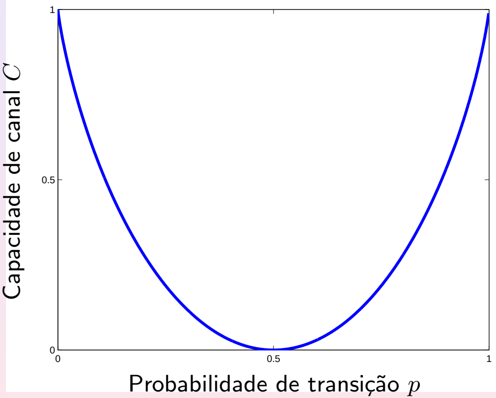
Canal Binário Simétrico (BSC)

- $\Pr(y_0|x_1) = \Pr(y_1|x_0) = p$
- $\Pr(y_0|x_0) = \Pr(y_1|x_1) = 1 - p$
- Sabendo que

$$\mathcal{I}(X,Y) = \sum_{k=0}^{K-1} \sum_{j=0}^{K-1} P(x_j, y_k) \cdot \log \left[\frac{P(y_k|x_j)}{P(y_k)} \right]$$

- Logo

$$\begin{aligned} C &= 1 + p \cdot \log[p] + (1-p) \cdot \log[1-p] \\ &= 1 - \mathcal{H}(p) \end{aligned}$$

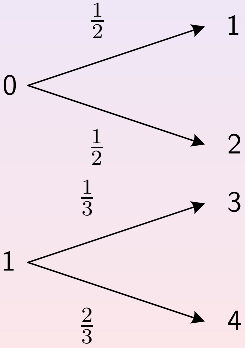

1
0.5
0
0
0.5
1
Capacidade de canal C
Probabilidade de transição p

- Curva simétrica (esperado!)
- Quando não há ruído ($p = 0$) a capacidade de canal é máxima, com um bit por uso do canal. Neste mesmo valor, a entropia atinge seu valor mínimo
- Com a probabilidade de erro $p = 1/2$, a capacidade atinge o menor valor enquanto a entropia atinge o máximo. Nestes casos, o canal é dito ser **sem utilidade** (*useless*)

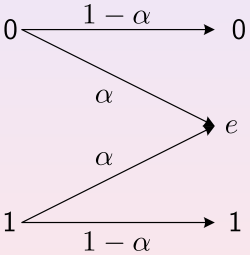
Capacidade de canal - outros exemplos

Canal com ruído com saídas não superpostas (não é canal com ruído!)

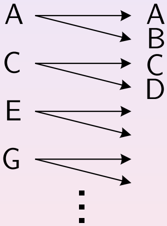
$$C = \max \mathcal{I}(X, Y) = 1 \text{ bit, para } p(x) = (\frac{1}{2}, \frac{1}{2})$$



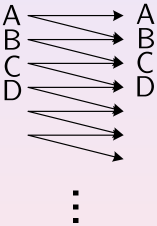
$$C = \max \mathcal{I}(X,Y) = 1 - \alpha \text{ bits}$$



$$C = \max \mathcal{I}(X, Y) = \log[13] \text{ bits}$$



Saídas não ruidosas



Canal com ruído

Codificação de canal - motivação

- Como melhorar a resistência do canal aos efeitos do ruído?
- Definição do limite de *confiabilidade*
- Idéia:
  - Transformar a seqüência de dados em uma seqüência de entrada do canal
  - Minimizar os efeitos do ruído (interferência)
  - Mapeamento inverso após o canal para recuperar seqüência de dados

Como fazer isso?

Inserção de **redundância**!

Codificação de canal - definições

- Simplificado para *códigos de bloco*
- Mensagem é subdividida em blocos seqüenciais cada um com k bits, e cada bloco é mapeado em um block de n bits em que $n > k$
- Número de bits de redundância: $n - k$
- A **taxa de código** é denominada a razão

$$r = \frac{k}{n} \tag{91}$$

- Acuracidade da recuperação requer probabilidade de erro tão pequena quanto possível

Questão importante

Existe algum método sofisticado de codificação de canal tal que a probabilidade de erro de seja menor que um número arbitrário positivo ϵ, em caso afirmativo, o esquema de codificação é eficiente tal que a taxa de codificação não necessite ser muito pequena?

Resposta

SIM!

Codificação de canal - definições

- A resposta é dada pelo teorema da capacidade de canal
- Generalização necessita de incluir o *tempo* como uma variável a ser considerada
- Suposições: fonte com alfabeto $\mathcal{A}$ e entropia $\mathcal{H}(A)$, emitindo sinais a cada T_s segundos
- Taxa média de informação

$$R = \frac{\mathcal{H}(A)}{T_s} \tag{92}$$

Codificação de canal - definições

- O decodificador tem de fornecer também símbolos do alfabeto $\mathcal{A}$ a uma taxa de T_s segundos
- Seja a capacidade do canal dada por C bits por transmissão
- Supondo que o canal é capaz de ser utilizando uma vez a cada T_c segundos
- Então, a *capacidade do canal por unidade de tempo*, representa a máxima taxa de transferência de informação pelo canal

$$\frac{C}{T_c} \tag{93}$$

Primeira parte

Seja uma fonte discreta com alfabeto $\mathcal{A}$ e entropia $\mathcal{H}(A)$ que produz símbolos a cada T_s segundos. Seja um canal discreto sem memória com capacidade de canal C e utilizado a cada T_c segundos. Então, se

$$\frac{\mathcal{H}(A)}{T_s} \leq \frac{C}{T_c} \tag{94}$$

existe um esquema de codificação para o qual a saída da fonte pode ser transmitida pelo canal e ser reconstruída com a uma probabilidade de erro arbitrariamente pequena. O parâmetro C/T_c é chamado de taxa crítica.

Segunda parte

Inversamente, se

$$\frac{\mathcal{H}(A)}{T_s} > \frac{C}{T_c} \tag{95}$$

não é possível transmitir informação pelo canal e reconstruí-la com uma probabilidade de erro arbitrariamente pequena.

Teorema da Codificação de Canal - discussão

- O teorema da codificação de canal é o resultado isolado mais importante da teoria da informação
- O teorema especifica que a capacidade de canal C é um *limite fundamental* na taxa na qual a transmissão é confiável sem erros, através de um canal discreto sem memória
- O teorema não mostra como construir um "bom código"
- Resultado: se a Equação (94) é satisfeita então **um** bom código existe

- Fonte: símbolos equiprováveis a cada T_s segundos, $\mathcal{H}(A) = 1$ bit por símbolo
- Taxa de informação da fonte: $1/T_s$
- Codificar de canal com taxa de codificação r produzindo símbolos a cada T_c segundos
- Taxa de transmissão de símbolos codificados é $1/T_c$
- A capacidade de canal por unidade de tempo é C/T_c bits por segundo
- Logo, se

$$\frac{1}{T_s} \leq \frac{C}{T_c} \tag{96}$$

a probabilidade de erro pode ser tão pequena quanto se deseje, usando uma codificação adequada

- A taxa de codificação é igual à razão

$$r = T_c/T_s \tag{97}$$

- Então, pode-se escrever

$$r \leq C \tag{98}$$

- Interpretação: para $r \leq C$ existe um código (com taxa de codificação menor ou igual a capacidade de canal) capaz de atingir uma taxa de erro arbitrariamente pequena

Teorema da capacidade de informação

Pergunta

Qual a capacidade de informação a ser transmitida por meio de um canal limitado em faixa, limitado em potência e gaussiano?

- Seja $X(t)$
  - Processo estocástico estacionário
  - Média nula
  - Faixa limitada em B Hz
- As K amostras discretas amostradas na taxa de Nyquist ($2B$ símbolos por segundo) são $X_k, k = 1, 2, \ldots, K$
- Amostras transmitidas a cada T segundos
- Canal limitado em faixa B Hertz
- Número de amostras

$$K = 2BT \tag{99}$$

Teorema da capacidade de informação

- Canal é a inserção de ruído aditivo gaussiano branco de média zero e densidade espectral de potência $N_0/2$
- Ruído limitado em B Hertz
- Logo

$$Y_k = X_k + N_k, \qquad k = 1, 2, \ldots, K \tag{100}$$

- Amostras de ruído são gaussianas, estatisticamente independentes, com média zero e variância

$$\sigma^2 = N_0 B \tag{101}$$

Teorema da capacidade de informação

- Como o canal é limitado em potência, cada entrada do canal deve ter uma *função custo*

$$\mathbb{E}\left\{X_k^2\right\} = P, \qquad k = 1, 2, \ldots, K \tag{102}$$

  em que P é a potência média de transmissão
- **Capacidade de informação** é definida como o máximo da informação mútua entre a entrada X_k e a saída Y_k sobre todas as distribuições de probabilidade de X_k que satisfazem a restrição da Equação (102)
- Matematicamente

$$C = \max_{p_X(x)} \left\{\mathcal{I}(X_k, Y_k) : \mathbb{E}\left\{X_k^2\right\} = P\right\} \tag{103}$$

Teorema da capacidade de informação

- Sabendo
$$\mathcal{I}(X_k, Y_k) = \mathcal{H}(Y_k) - \mathcal{H}(Y_k|X_k) \tag{104}$$
- E como X_k e N_k são independentes e a soma é igual a Y_k, tem-se
$$\mathcal{I}(X_k, Y_k) = \mathcal{H}(Y_k) - \mathcal{H}(N_k) \tag{105}$$
- Como $\mathcal{H}(N_k)$ é independente da distribuição de X_k, maximizar a Equação (105) é maximizar a entropia de Y_k
- Para uma variância fixa, Y_k terá máxima entropia se $p_Y(y)$ for uma distribuição gaussiana (será provado mais adiante!)
- Logo, X_k também dever ser gaussiano e atender à restrição de potência

- Então, escreve-se

$$C = \mathcal{I}(X_k, Y_k) : X_k \text{ gaussiano}, \ \mathbb{E}\left\{X_k^2\right\} = P \qquad (106)$$

- Cálculos
  1. Variância de Y_k é $P + \sigma^2$, logo

$$\mathcal{H}(Y_k) = \frac{1}{2} \log\left[2\pi e(P + \sigma^2)\right] \qquad (107)$$

  2. Variância do ruído N_k é σ^2, assim

$$\mathcal{H}(N_k) = \frac{1}{2} \log\left[2\pi e \sigma^2\right] \qquad (108)$$

- Substituindo as Equações (107) e (108) na Equação (105) tem-se

$$C = \frac{1}{2} \log \left(1 + \frac{P}{\sigma^2} \right) \text{ bits por transmissão} \tag{109}$$

- Lembrando que o canal é usado K vezes para transmitir as K amostras em T segundos, tem-se que a informação é dada por *capacidade por unidade de tempo* (K/T). Também $K = 2BT$, logo

$$C = B \log \left(1 + \frac{P}{N_0 B} \right) \text{ bits por segundo} \tag{110}$$

Teorema da capacidade de informação

Terceiro teorema de Shannon

A capacidade de informação de um canal contínuo de largura de faixa B Hertz corrompido com ruído aditivo gaussiano branco de densidade espectral de potência $N_0/2$ e limitado em faixa em B Hertz é dado por

$$C = B \log\left(1 + \frac{P}{N_0 B}\right) \text{ bits por segundo} \tag{111}$$

- Resultado mais famoso de Shannon
- Interrelaciona os parâmetros chaves de um sistema de transmissão: largura de faixa do canal, potência média transmitida e densidade espectral de potência
- O teorema implica que, para uma dada P e largura de faixa do canal B pode-se transmitir informação a uma taxa de C bits por segundo com uma *probabilidade de erro arbitrariamente pequena*, utilizando suficientemente complexos sistemas de codificação
- **Não** é possível transmitir a uma taxa maior que C bits por segundo usando nenhum sistema de codificação e ter uma probabilidade de erro arbitrariamente pequena
- *Limite fundamental* da transmissão sem erros
- Para se aproximar deste limite o sinal deve ter estatísticas próximas do ruído gaussiano branco

Implicações do teorema da capacidade de informação

- Sendo o canal limitado em faixa e potência, qual o impacto do teorema da capacidade de informação?
- *Framework* prático - sistema ideal
- Taxa de bits R_b igual à capacidade de informação C
- Potência média de transmissão

$$P = E_b C \tag{112}$$

em que E_b é a **energia por bit**

- Logo, para o sistema ideal

$$\frac{C}{B} = \log\left(1 + \frac{E_b}{N_0}\frac{C}{B}\right) \tag{113}$$

Implicações do teorema da capacidade de informação

- De maneira equivalente pode-se definir a relação **energia do sinal por bit pela densidade espectral de potência** em termos da eficiência de faixa como

$$\frac{E_b}{N_0} = \frac{2^{C/B} - 1}{C/B} \tag{114}$$

- Eficiência de faixa $\frac{C}{B}$

Implicações do teorema da capacidade de informação
Curva de eficiência de faixa - limite de Shannon