Teoria da Informação e Codificação

Jorge Almeida

Gabinete 3.33

jalmeida@fc.up.pt

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO
Departamento de Matemática

U.PORTO

Apresentação

- Programa preliminar
- Bibliografia
- Avaliação

Programa preliminar

- Informação e entropia
- Canais de informação
- Codificação da fonte
- Compressão de dados
- Codificação do canal
- Códigos corretores de erros

Bibliografia

[1] Robert B. Ash, *Information Theory*, Dover, New York, 1990.

[2] Gareth A. Jones and J. Mary Jones, *Information and Coding Theory*, Springer, London, 2000.

[3] Steven Roman, *Introduction to coding and information theory*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1997.

[4] Roberto Togneri and Christopher J. S. deSilva, *Fundamentals of Information Theory and Coding design*, Chapman & Hall/CRC, Boca Raton, 2003.

Avaliação

- Teste e exame, sendo o resultado final a média das duas notas.
- Todos os alunos inscritos têm acesso às provas de avaliação.
- Na época de recurso e nas épocas especiais, a nota é aquela que for obtida no exame.

O que é a teoria da informação?

- A teoria da informação nasceu nos finais da década de 1940 com os trabalhos de Claude E. Shannon, nomeadamente com o artigo *A mathematical theory of communication*, Bell System Technical Journal **27** (1948), 379–423, 623–656.
- Os modelos introduzidos por Shannon visavam a resolução de um problema que na altura parecia intratável: como transmitir informação, eficientemente e de forma correta, através de canais que aleatoriamente a modificam?

Há efetivamente dois problemas na transmissão de informação:

- informação como texto, som, imagem, vídeo, para o seu fácil reconhecimento pelos seres humanos contém em geral grande redundância;
  - por exemplo, em português, com grande probabilidade, a seguir a um ç encontramos um dos ditongos ão ou ões;
  - no vídeo, como sequência de imagens, há em geral uma correlação muito grande entre imagens consecutivas;

- devido a interferências inerentes à natureza dos canais de comunicação, há perda de informação na sua utilização; por exemplo,
  - sinais eléctricos são sujeitos a interferências eletromagnéticas,
  - sinais de rádio via satélite sofrem interferências de radiações cósmicas,
  - e ambos recebem interferências de outros sinais do mesmo tipo.

Ideias óbvias

Há duas ideias óbvias para atacar estes problemas:

- reduzir a redundância na fonte de informação para poupar (em tempo/custos reais) na utilização dos canais de comunicação;
  $\rightarrow$ compressão dos dados
- acrescentar criteriosamente redundância na informação efetivamente transmitida de forma a que o recetor possa recuperar a mensagem inicial daquela que lhe chega contendo erros resultantes da passagem pelo canal de comunicação.
  $\rightarrow$ expansão da mensagem

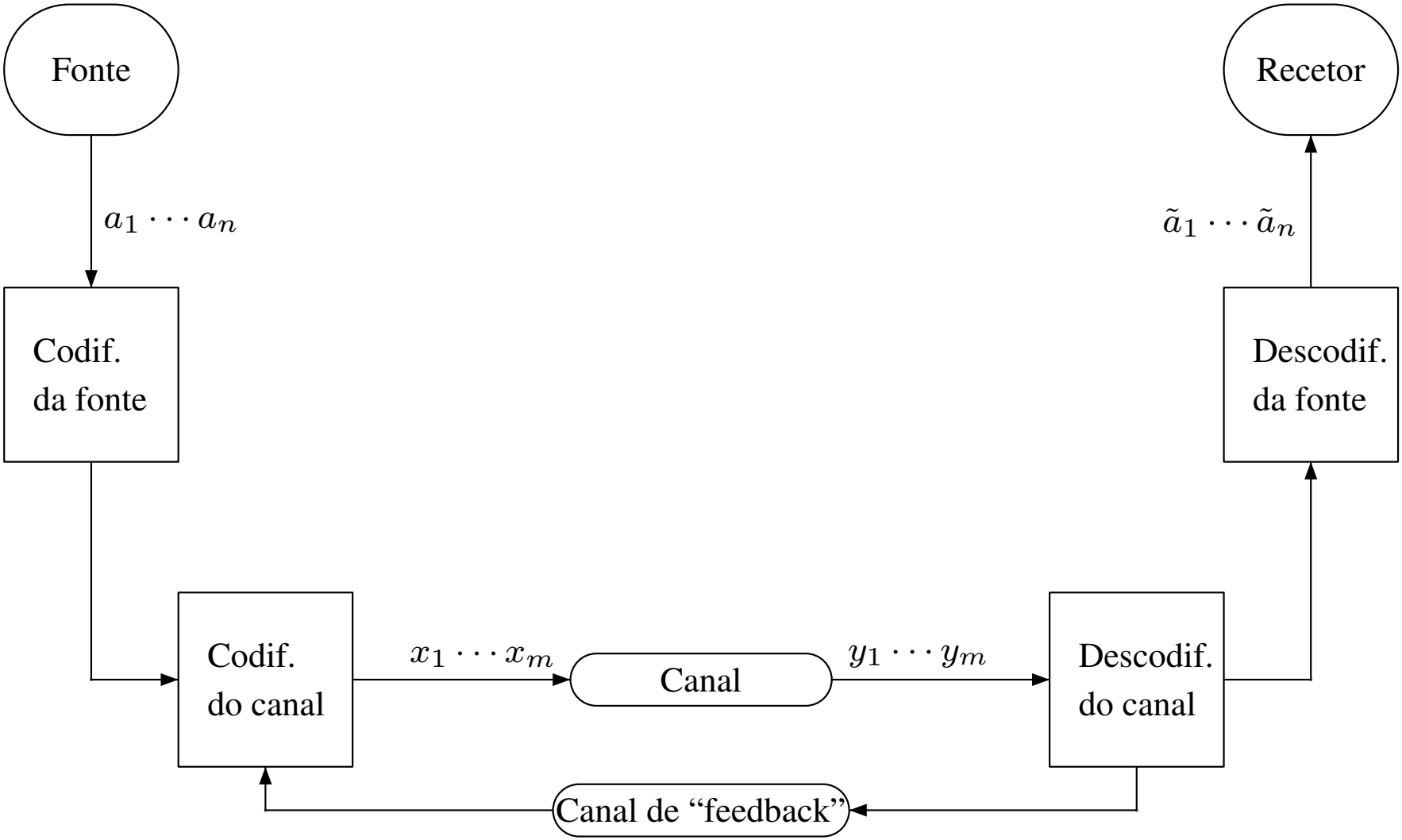
Dificuldades

- A grande dificuldade está em saber em que medida estas ideias podem resultar:
  - é de esperar que haja limitações inerentes à própria natureza do canal de comunicação;
  - até onde é que será necessário/possível ir com estas técnicas?

As ideias de Shannon

- As contribuições seminais de Shannon são:
  - quantificação dos parâmetros envolvidos (fonte de informação e perturbações no canal de comunicação) usando métodos probabilísticos;
    $\rightarrow$ noção de entropia em informação
  - separação dos dois problemas, nomeadamente a codificação da fonte e a codificação do canal;
  - estabelecimento de limites teóricos para a resolução do problema;
  - prova da existência de aproximações arbitrariamente boas dos limites teóricos.
    $\rightarrow$ métodos para a compressão de dados e
    $\rightarrow$ códigos corretores de erros

Esquema da comunicação fonte $\rightarrow$ recetor



1 Informação e entropia

- Distribuição de probabilidade
- Entropia
- Unidades de entropia
- Valores extremos da entropia

Distribuição de probabilidade (discreta)

- Uma *distribuição de probabilidade* num espaço de amostragem $S = \{s_1, \ldots, s_N\}$ é uma função $P : S \to [0, 1]$ tal que $\sum_{s \in S} P(s) = 1$.
- Um *evento* é um subconjunto E do espaço de amostragem S.
- A *probabilidade* de um evento E é $P(E) = \sum_{s \in E} P(s)$.
- Note-se que $P(\emptyset) = 0$ e, para eventos E e F, tem-se

$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

Valor esperado e surpresa

- Dada uma função $f : S \to V$ com valores num espaço vetorial real V, o *valor esperado* de f é a média pesada dos valores de f dada por

$$\bar{f} = \sum_{s \in S} P(s) f(s).$$

- Face a uma certa distribuição da probabilidade, a ocorrência de um certo evento pode ser mais ou menos surpreendente.
  Definimos a *surpresa* do evento E como sendo

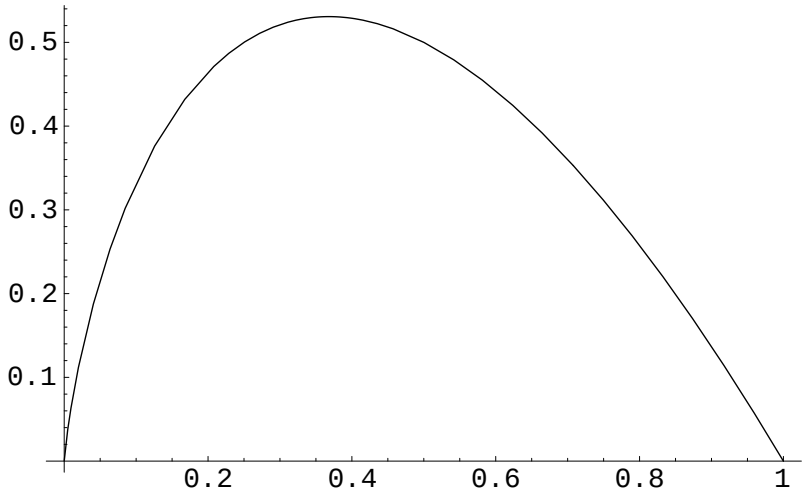
$$\mathrm{surp}(E) = -\log(P(E)) = \log(1/P(E)).$$

Entropia

- A *entropia* da distribuição de probabilidade P é o valor esperado da surpresa (restrita a subconjuntos singulares de S):

$$H(P) = -\sum_{s \in S} P(s) \log(P(s)).$$

- Gráfico da função $-p \log(p)$ (para o logaritmo na base 2):



o máximo é atingido para $p = 1/e$, seja qual for a base do logaritmo.

Entropia em termodinâmica

- A noção de entropia já tinha sido introduzida no séc. XIX no âmbito da termodinâmica, onde entropia é uma medida da desordem de um sistema.
- A *Segunda Lei da Termodinâmica* afirma que um sistema (termodinâmico) não pode, sem intervenção exterior, diminuir a sua entropia. Ou seja, a tendência natural dos sistemas é para aumentar a sua entropia.

Exemplos de cálculo da entropia (1)

- Seja $S = \{s_1, s_2\}$ com $P(s_1) = P(s_2) = 0.5$. Então

$$H(P) = -2(1/2)\log(1/2) = \log(2) = 1.$$

  Aqui há desordem completa, sendo completamente imprevisível o acontecimento a observar.

- Seja $S = \{s_1, s_2\}$ com $P(s_1) = 0.98$ e $P(s_2) = 0.02$. Então

$$H(P) \simeq -(0.98)(-0.0291463) - (0.02)(-5.64386) \simeq 0.14.$$

  Aqui, a surpresa está concentrada na ocorrência de um acontecimento, muito pouco provável, ao que corresponde um valor muito menor da entropia.

Exemplos de cálculo da entropia (2)

- Seja $S = \{s_1, s_2\}$ com $P(s_1) = 1$ e $P(s_2) = 0$. Então, tomando $0 \log(0) = \lim_{x \to 0^+} x \log(x) = 0$, temos

$$H(P) = -(1) \log(1) - (0) \log(0) = 0.$$

  Obtemos entropia nula, não havendo qualquer lugar para surpresa, o acontecimento a observar é garantidamente s_1.

Exemplos de cálculo da entropia (3)

- Seja $S = \{s_1, \ldots, s_6\}$ com $P(s_i) = 1/6$ $(i = 1, \ldots, 6)$. A entropia é $H(P) = \log(6) \simeq 2.585$, sendo novamente a desordem completa.
- Seja $S = \{s_1, \ldots, s_6\}$ com $P(s_i) = 0.498$ $(i = 1, 2)$ e $P(s_j) = 0.001$ $(j = 3, \ldots, 6)$. A entropia é

$$\begin{aligned} H(P) &= -2 * 0.498 * \log(0.498) - 4 * 0.001 * \log(0.001) \\ &\simeq 1.0412 \end{aligned}$$

  ou seja um valor próximo do primeiro exemplo, em que havia só dois acontecimentos igualmente prováveis.

Unidade da entropia

- O valor da entropia depende da base do logaritmo considerada.
- Normalmente tomamos logaritmos na base 2 e chamamos *bit* (*b*inary un*it*) à correspondente unidade de entropia. Se a base considerada for a base e do logaritmo natural, falamos de *nits* (*n*atural un*it*).
- A mudança de unidade de entropia faz-se de acordo com a fórmula

$$H_r = \frac{H_e}{\ln(r)},$$

onde H_x representa a entropia para logaritmos na base x.

Uma desigualdade útil

Lema 1.1 *Sejam $p_1, \ldots, p_N$ e $q_1, \ldots, q_N$ números reais não negativos tais que $\sum_{i=1}^{N} p_i = \sum_{i=1}^{N} q_i = 1$. Então*

$$-\sum_{i=1}^{N} p_i \log(p_i) \leqslant -\sum_{i=1}^{N} p_i \log(q_i) \tag{1}$$

e a igualdade verifica-se sse $p_i = q_i$ para todo o i.

Prova. Basta considerar o caso do logaritmo natural pois $\log_r(x) = \ln(x)/\ln(r)$. Note-se que $\ln x \leqslant x - 1$ para todo o $x > 0$, com igualdade sse $x = 1$. Podemos ainda ignorar os termos em que $p_i = 0$ ou $q_i = 0$.
Logo tem-se $\ln(q_i/p_i) \leqslant q_i/p_i - 1$ com igualdade sse $p_i = q_i$. Multiplicando por p_i e somando, obtém-se

$$\sum_{i=1}^{N} p_i \ln(q_i/p_i) \leqslant \sum_{i=1}^{N} (q_i - p_i) = 1 - 1 = 0.$$

Se a igualdade entre as somas se verificar, sendo a desigualdade válida termo a termo, então $p_i = q_i$ para todo o i. □

Valores extremos da entropia

Teorema 1.2 *Se o espaço de amostragem tem N elementos, então os valores extremos para a entropia, como função da distribuição de probabilidade $P = (p_1, \ldots, p_N)$ são:*

- *mínimo:* 0*, atingido exactamente quando um dos* $p_i = 1$*;*
- *máximo:* $\log(N)$*, atingido exactamente quando todos os* $p_i = 1/N$*.*

Prova. Que aqueles valores, 0 e $\log(N)$ correspondem às distribuições de probabilidade indicadas segue por cálculo direto.

Que 0 é o valor mínimo é óbvio pois, por definição, a entropia é $\geqslant 0$. Também da definição segue que só aquelas distribuições de probabilidade têm entropia 0.

Que $\log(N)$ é o valor máximo segue do Lema 1.1 tomando $q_i = 1/N$. A condição para a igualdade do Lema mostra que o máximo só é atingido para a distribuição uniforme. □