



Teoria da Informação - TIP 812

Prof. Dr. Charles Casimiro Cavalcante

Número de créditos: 4

Carga horária total: 60 h

Período: 2010.1

Lista de Exercícios No. 3: Codificação de Fontes

- Seja X uma variável aleatória com n valores possíveis $x_1, x_2, \dots, x_{n-1}, x_n$ de probabilidades $\frac{1}{2}, \frac{1}{2^2}, \dots, \frac{1}{2^{n-1}}, \frac{1}{2^n}$.
 - Implemente uma codificação de Huffman dos n possíveis valores de X ;
 - Compare o comprimento médio das palavras códigos L com a entropia $\mathcal{H}(X)$ da fonte. O que se nota deste resultado? Em que este resultado é importante? Como explica-se este resultado?
- Considere uma fonte binária sem memória A cuja distribuição de probabilidade é $\Pr(A = 0) = 0,9$ e $\Pr(A = 1) = 0,1$. Como os zeros são muito mais frequentes que os uns, propõe-se de codificar as seqüências emitidas da fonte levando-se em conta o número de zeros entre dois uns consecutivos. A operação consiste então de duas etapas:
 - Primeira etapa:
Conta-se o número de zeros entre dois uns consecutivos. Obtém-se assim um inteiro denominado *inteiro intermediário*.
 - Segunda etapa:
Codifica-se o inteiro intermediário em uma palavra binária constituída de quatro elementos binários se o inteiro é inferior ou igual a 7 e escolhe-se uma palavra de um elemento binário se o inteiro é 8. Se o inteiro intermediário ultrapassa 8, codifica-se as seqüências de 8 zeros consecutivos por um bit correspondente ao inteiro intermediário quantas vezes for necessário. Desta forma, obtém-se a seguinte tabela de correspondência entre as seqüências da fonte e os inteiros intermediários.

Seqüência da fonte	Inteiro intermediário
1	0
01	1
001	2
0001	3
...	...
...	...
00000001	7
00000000	8

Exemplo relativo a primeira etapa da atribuição dos inteiros intermediários:

1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 1
0 2 8 2 0 4

- As restrições impostas permitem escolher um código unicamente decodificável?
- Calcular o número médio de bits L_1 da fonte por inteiro intermediário.
- Calcular o número médio de bits L_2 das palavras codificadas por inteiro intermediário.



- (d) Efetuar uma codificação de Huffman de ordem 4 da fonte A . Calcular o número médio de bits utilizados para codificar um elemento binário emitido pela fonte.

3. Seja uma fonte discreta sem memória de valores possíveis $\{a_1, a_2, \dots, a_i, \dots\}$ que verifique:

$$\forall k > j \geq 1, \quad \Pr(a_k) \leq \Pr(a_j)$$

Propõe-se de codificar os diferentes símbolos da fonte A da seguinte maneira:

Para o símbolo a_i é associado um número Q_i definido por $Q_i = \sum_{k=1}^{i-1} \Pr(a_k) \forall k > 1$ e $Q_1 = 0$.

A palavra código binária c_i correspondente ao símbolo a_i é obtida tomando-se a parte “decimal” do valor de Q_i no sistema binário (por exemplo, $Q_i = \frac{1}{2} \rightarrow 100\dots$, $Q_i = \frac{1}{4} \rightarrow 0100\dots$, $Q_i = \frac{5}{8} \rightarrow 10100\dots$) que nós truncamos sobre o número de bits igual ao menor inteiro superior ou igual a auto-informação do evento a_i .

- (a) O código C verifica a condição de prefixo (inequação de Kraft-McMillan)?
- (b) Denota-se por L o comprimento médio das palavras código. Mostrar que L verifica a inequação $\mathcal{H}(A) \leq L \leq \mathcal{H}(A) + 1$.
- (c) Codificar pelo procedimento descrito os oito símbolos de uma fonte sem memória de probabilidades $\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}$
- Em relação ao item anterior, calcular o comprimento médio das palavras código obtidas e comparar com a entropia da fonte. O que pode-se constatar? Como explicar este resultado?

4. Utilizar a codificação de Lempel-Ziv para codificar a sequência contida no arquivo no seguinte *site*:

http://www.gtcl.ufc.br/~charles/PDF/it/LZ_data.txt

Dica: Use o MatLab para carregar o arquivo e visualizar os dados. Deve ser uma sequência de '0's e '1's.

- (a) Qual a taxa de codificação?
- (b) Quando o código seria mais eficiente?