

Exame de Compressão e Codificação de Dados

Secção de Telecomunicações – DEEC
Instituto Superior Técnico

10 de Fevereiro de 2000

1. **a)** Considere uma fonte que emite símbolos do alfabeto $\mathcal{X} = \{1, 2, 3, 4, 5\}$. Determine (justificando) uma distribuição de probabilidades $\{p(1), \dots, p(5)\}$ por forma a que seja possível construir para esta fonte um código cujo comprimento média seja igual à entropia. Desenhe um código de Huffman binário (para essa distribuição de probabilidades) e calcule o seu comprimento médio (caso não tenha conseguido determinar as probabilidades pretendidas, considere $p(1) = 0.4$, $p(2) = 0.2$, $p(3) = 0.15$, $p(4) = 0.15$, $p(5) = 0.1$).
- b)** Considere a extensão de segunda ordem da fonte (com as probabilidades determinadas na alínea anterior) e o correspondente código de Huffman binário. Qual o ganho em eficiência (em bits/símbolo)?
- c)** A fonte X gera aleatoriamente números pertencentes ao conjunto $\{0, 1, 2, \dots, 20\}$. Qual a relação ($>$, $<$, ou $=$) entre a entropia de X e as entropias de $Y = X^2$ e $Z = (X - 10)^2$? Justifique. Sabendo que a entropia de X é $\log_2 21$ bits/símbolo, determine as informações mútuas $I(X; Y)$ e $I(X; Z)$.
- d)** Existe um procedimento de codificação (proposto por Tunstall, em 1968) que pode ser visto como recíproco do de Huffman. No código de Huffman procura-se codificar os símbolos da fonte com sequências de comprimento variável por forma a que aos símbolos mais prováveis correspondam sequências mais curtas. Na codificação de Tunstall, o objectivo é codificar sequências (com comprimento variável) de símbolos da fonte com palavras de código de comprimento fixo. Por exemplo, um código de Tunstall binário para uma fonte que emite símbolos $\{A, B\}$ pode ser algo como

$$AAA \rightarrow 00, \quad AAB \rightarrow 01, \quad AB \rightarrow 10, \quad B \rightarrow 11$$

Note que, dado que o código de Tunstall tem comprimento fixo, é um código instantâneo. No procedimento para construção deste código também se constrói uma árvore binária, mas de forma inversa à de Huffman: em vez de se combinarem os símbolos menos prováveis, decompõe-se a sequência mais provável em duas, até que o total de folhas da árvore atinga um número predefinido 2^n , em que n é o comprimento das palavras do código de Tunstall ($n = 2$, no exemplo acima). O objectivo natural é maximizar o comprimento médio das sequências de entrada do código. Construa um código de Tunstall de comprimento $n = 2$ para uma fonte de símbolos $\{A, B\}$ com $p(A) = 0.8$. Calcule a eficiência média do código obtido (em bits/símbolo); se não conseguiu obter o código, considere o indicado no exemplo acima. Desenhe um código de Huffman para a extensão de ordem 2 da fonte (isto é, para $\{AA, AB, BA, BB\}$) e compare os resultados.

2. Uma fonte de Markov de primeira ordem é caracterizada pela distribuição de probabilidade do símbolo X_n , condicionada ao símbolo X_{n-1} , designada por $p(X_n|X_{n-1})$. Quando a fonte é estacionária, esta distribuição não depende de n , pelo que $p(X_n|X_{n-1}) = p(X_m|X_{m-1})$, para quaisquer dois instantes n e m . Considere então uma fonte de Markov, gerando símbolos do alfabeto $\{A, B, C\}$, cuja distribuição de probabilidade condicionada é

$p(X_n X_{n-1})$	$X_n = A$	$X_n = B$	$X_n = C$
$X_{n-1} = A$	1/3	1/3	1/3
$X_{n-1} = B$	1/4	1/2	1/4
$X_{n-1} = C$	0	1	0

Sabe-se que $\mu = [0.2143, 0.5714, 0.2143]$ é a distribuição estacionária desta fonte de Markov, isto é que $\mu = \mu\mathbf{A}$, em que $\mathbf{A}$ é a matriz de definida na tabela anterior.

- a) Recorde que um processo possui taxa de entropia condicional, dada por $\lim_{n \rightarrow \infty} H(X_n | X_{n-1})$, se este limite existir. Calcule a taxa de entropia condicional deste processo. Compare com a entropia da distribuição estacionária e comente.
 - b) Suponha agora que se pretende desenhar um sistema de codificação para esta fonte e que se opta por desenhar 3 códigos diferentes, um para cada possível símbolo anterior. Desenhe estes códigos e calcule o comprimento médio de cada um deles, condicionado ao respectivo símbolo anterior. Com base no vector de distribuição estacionário acima indicado, calcule o comprimento médio geral deste procedimento de codificação; compare com a taxa de entropia e com a entropia da distribuição estacionária e comente.
 - c) Desenhe agora um código de Huffman para a extensão de segunda ordem desta fonte, calcule o comprimento médio, e compare com o resultado da alínea anterior.
3. Considere o seguinte procedimento de codificação do tipo Lempel-Ziv: cada subsequência de símbolos de comprimento superior ou igual a L que já tenha ocorrido anteriormente é substituída por um par de “ponteiros” com o formato (p, l) , indicando que a ocorrência anterior se verificou na p posições atrás e que o comprimento da subsequência é l . O código obtido é binário, sendo cada símbolo (não substituído) representado pelo respectivo código ASCII (em 8 bits); por exemplo, ASCII('A') = 01000001 e ASCII('D') = 01000100. Por sua vez, os ponteiros são também escritos em binário, com 16 bits para p e 8 bits para l . Assim, por exemplo, com $L = 2$, a sequência DADADADA é codificada como (os espaços servem apenas para tornar mais clara a leitura, não fazendo parte do código)

01000100 01000001 0000000000000010 00000110

- a) Justifique a seguinte afirmação: o esquema apresentado não produz um código descodificável. Justifique, mostrando que não é possível descodificar a sequência binária acima por forma a obter DADADA.
 - b) Sugira uma modificação do esquema por forma a obter códigos descodificáveis. Aplique o esquema sugerido, com $L = 2$, à sequência DADADA.
 - c) Para o novo esquema, dimensione L por forma a que a codificação nunca produza uma sequência de comprimento maior do que a original.
4. Considere um quantizador escalar uniforme, cuja gama de valores de entrada é $[-10, 10[$. Assuma que a entrada é uma variável aleatória X cuja função densidade de probabilidade é uniforme no intervalo $[-10, 10[$.
- a) Determine o número de células do quantizador por forma a que a relação sinal/ruído de quantização SNR_Q seja de 20dB. Recorde que $SNR_Q = 10 \log_{10}(\sigma_s^2 / \sigma_r^2)$, em que σ_s^2 e σ_r^2 são, respectivamente, a variância do sinal e do erro de quantização. Este quantizador é ótimo para X ? Existe distorção de sobrecarga (*overload*)?
 - b) Considere agora que a entrada tem uma função densidade de probabilidade uniforme no intervalo $[-12, 12[$. Qual é a probabilidade de sobrecarga (*overload*)? Qual é a variância do erro de sobrecarga?