

Teoria da Informação

Trabalho Prático Laboratorial nº1

Bruno José Borges Madureira 2011161942
Bruno Machado de Sousa Leão 2012138765

Introdução:

Neste trabalho foi nos dado um conjunto de dados ao qual tivemos de aplicar fórmulas matemáticas para a determinação da entropia, entropia conjunta e o número de bits necessários a codificação por códigos de huffman.

Entropia é o número médio de bits para codificar uma fonte de Informação.



The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.

Exercicio 1:



Para descobrir o numero de ocorrências de cada elemento do alfabeto na cadeia de símbolos de entrada, construímos um ciclo for em que iteramos cada elemento do alfabeto e contávamos o numero de ocorrências de cada elemento na cadeia de entrada.

Depois com o auxilio da função “bar” fizemos plot do numero de ocorrências de cada símbolo.



Exercicio 2:



The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.

O limite mínimo teórico para o numero médio de bits por símbolo é a entropia. Neste exercício recorreremos a 3 funções, uma que calcula a ocorrência de cada símbolo sem fazer plot (“Histograma_sem_grafico”), uma que calcula a probabilidade de cada simbolo (“probabilidade”) e outra que aplica a formula da entropia tal como foi apresentada nas aulas (“Entropia”).



The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.

Exercicio 3:



The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.

kid.bmp:



The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.

Entropia: 6.954

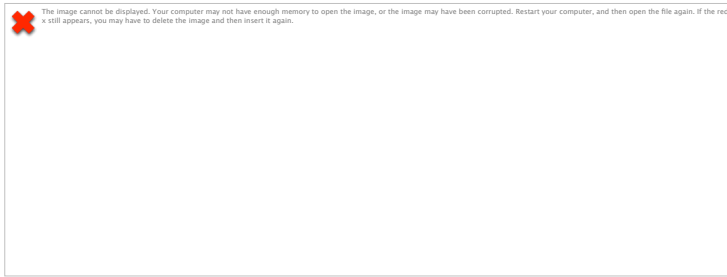
homer.bmp:



The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.

Entropia: 3.466

homerBin.bmp:

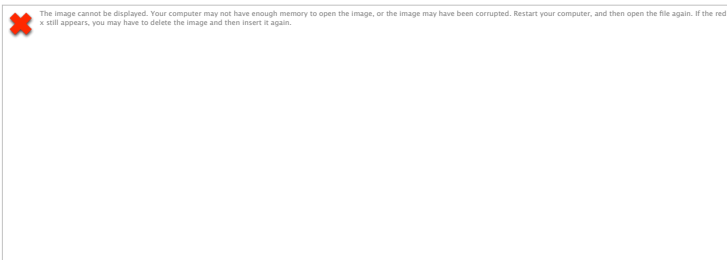


Entropia: 0.645

guitarSolo.wav:

Entropia: 7.358

english.txt:



Entropia: 4.228

Como era esperado, há maior entropia nos ficheiros onde há maior dispersão dos dados.

Como não estamos a descartar dados nenhuns em nossa análise, chegámos a conclusão de que é possível sim comprimir cada um dos ficheiros de forma não destrutiva, sendo a compressão máxima alcançável igual a $N \cdot M$, sendo N o número de símbolos da fonte e M a entropia da mesma.

Exercicio 4:



The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.

kid.bmp:

Hufflen: 6.983

homer.bmp:

Hufflen: 3.548

homerBin.bmp:

Hufflen: 1

guitarSolo.wav:

Hufflen: 7.379

english.txt:

Hufflen: 4.252

Como podemos observar o número de bits necessários para codificar nossas informações aproxima-se muito da entropia de cada fonte, obtemos uma variância maior entre os códigos de huffman e a entropia no arquivo homerBin, isso porque a entropia de tal arquivo é um valor muito abaixo de 1, e o código de huffman está adaptado para situações reais, logo, não pode codificar uma informação com menos de 1 bit.

Exercicio 5



The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.

kid.bmp:

Entropia Conjunta: 4.909

homer.bmp:

Entropia Conjunta: 2.413

homerBin.bmp:

Entropia Conjunta: 0.398

guitarSolo.bmp:

Entropia Conjunta: 5.781

english.txt:

Entropia Conjunta: 3.652

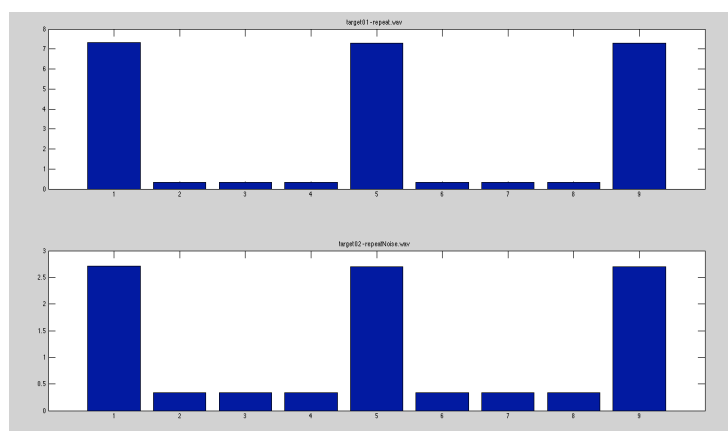
O valor da entropia conjunta é sempre menor ou igual que o valor da entropia por cada símbolo, então chegámos a conclusão que é sempre benéfico codificarmos os nossos dados usando agrupamentos de símbolos.

Exercicio 6a

- a) Escreva uma rotina em Matlab que, dada a query, o target, um alfabeto $A=\{a_1, \dots, a_n\}$ e o passo, devolva o vector de valores de informação mútua em cada janela.

Exercicio 6b

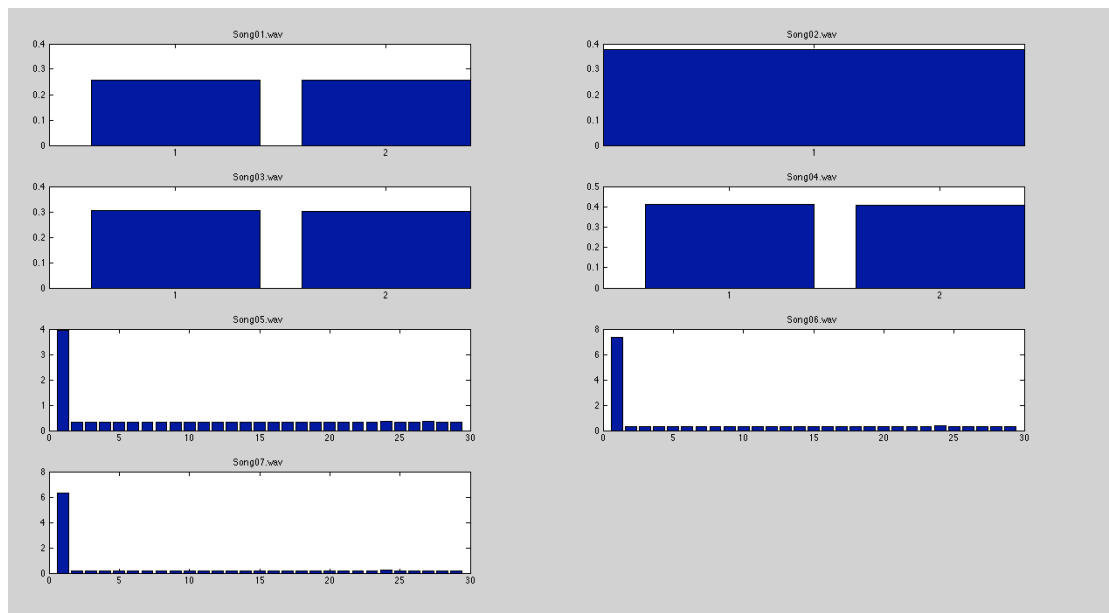
- b) Usando o ficheiro "guitarSolo.wav" como query, determine a variação da informação mútua entre este e os ficheiros "target01 - repeat.wav" e "target02 - repeatNoise.wav". Defina um passo com valor de $\frac{1}{4}$ do comprimento do vector da query (valor arredondado).



Quanto maior a semelhança audível, maior informação mútua, logo a variação da informação mútua no caso do target1 e target2 é perceptível ao olharmos o gráfico de cada.

Exercicio 6c

- c) Pretende-se agora simular um pequeno simulador de identificação de música. Usando o ficheiro "guitarSolo.wav" como query e os ficheiros Song*.wav como target:
- determine a evolução da informação mútua para cada um dos ficheiros
 - calcule a informação mútua máxima em cada um deles
 - e, finalmente, apresente os resultados da pesquisa, seriados por ordem decrescente de informação mútua. Defina um passo com valor de $\frac{1}{4}$ da duração da query.



Informação mútua entre guitarSolo.wav e Song06.wav: 7.338
Informação mútua entre guitarSolo.wav e Song07.wav: 6.313
Informação mútua entre guitarSolo.wav e Song05.wav: 3.962
Informação mútua entre guitarSolo.wav e Song04.wav: 0.410
Informação mútua entre guitarSolo.wav e Song02.wav: 0.378
Informação mútua entre guitarSolo.wav e Song03.wav: 0.304
Informação mútua entre guitarSolo.wav e Song01.wav: 0.258

Como era esperado, há maior informação mútua quando há maior semelhança audível nos ficheiros.

