Teoria

da

Informação

Trabalho Pratico 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.



Exercicio 1:



Para descobrir o numero de ocorrências de cada elemento do alfabeto na cadeia de símbolos de entrada, construi mos um ciclo for em que iteramos cada elemento do alfabeto e contava mos 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:



O limite mínimo teórico para o numero médio de bits por símbolo é a entropia. Neste exercício recorremos 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”).



Exercicio 3:



kid.bmp:



Entropia: 6.954

homer.bmp:

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\*M, sendo N o número de símbolos da fonte e M a entropia da mesma.

Exercicio 4:

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  
  


kid.bmp:

Entropia Conjunta: 4.909

homer.bmp:

Entropia Conjunta: 2.413

homerBin.bmp:

Entropia Conjunta: 0.398

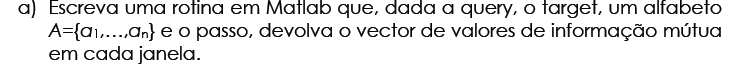
guitarSolo.bmp:

Entropia Conjunta: 5.781

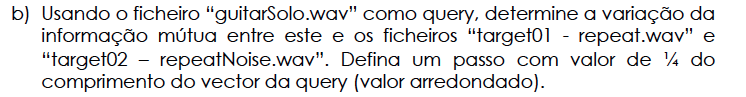
english.txt:

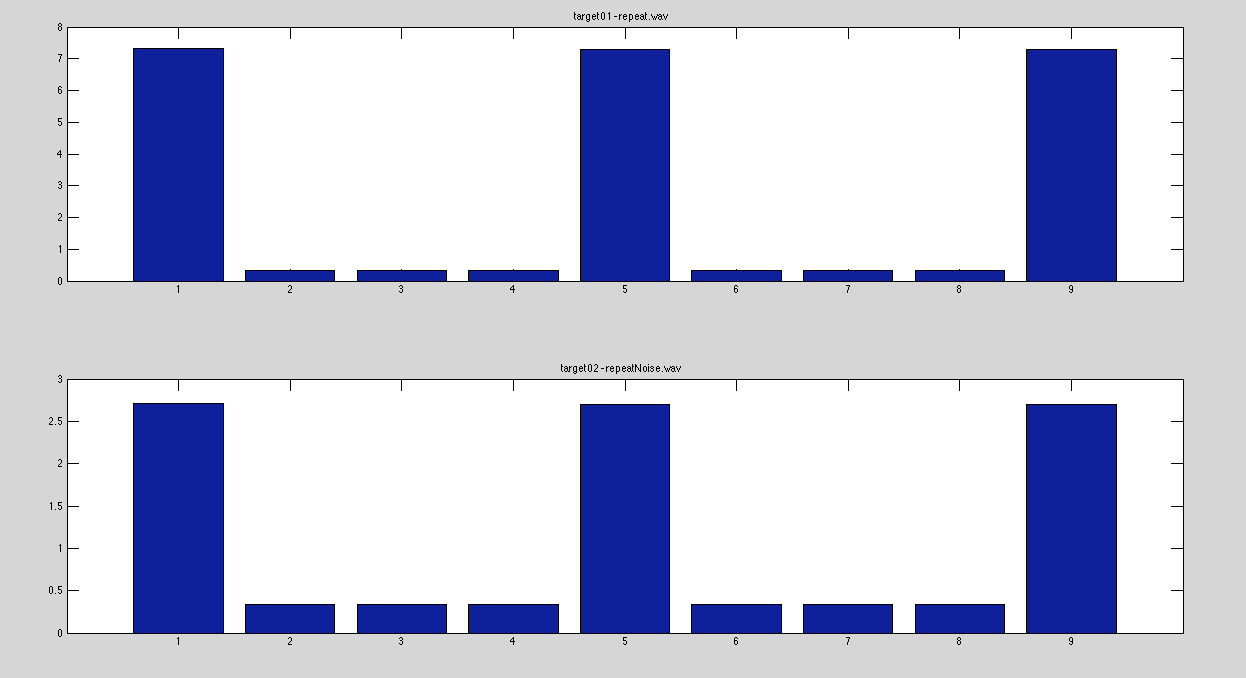
Entropia Conjunta: 3.652

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  
  


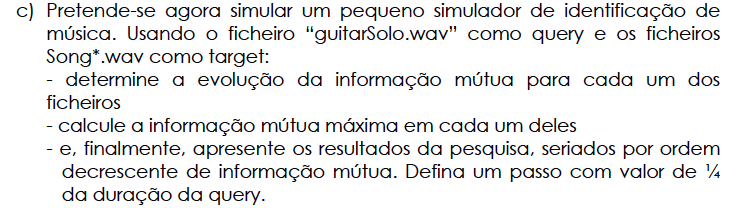
Exercicio 6b

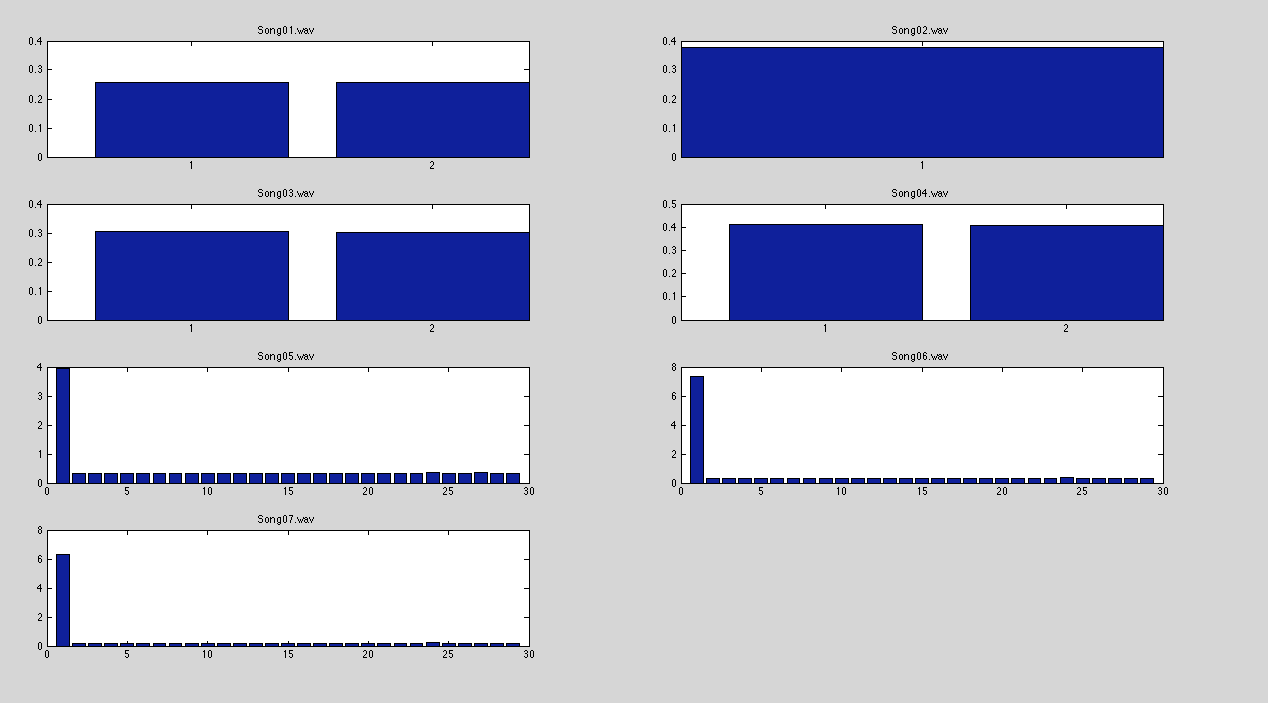




Quanto maior a semelhança audível, maior informação mútua, logo a variação da informação mútua no caso do target1 e target2 é perceptível ao olharmos o gráfico de cada.

Exercico 6c





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.