

Teoria da Informação

Relatório Trabalho Prático 1



Entropia, Redundância e Informação Mútua

Tiago Batista N°2014198201
João Mendes N°2016230975
João Montenegro N°2016228672

Exercício 3:

Resultados Obtidos:

kid.bmp: Entropia 6.9541 bits/símbolo

homer.bmp: Entropia 3.4659 bits/símbolo

homerBin.bmp: Entropia 0.6448 bits/símbolo

guitarSolo.wav: Entropia 7.3580 bits/símbolo

english.txt: Entropia 4.2280 bits/símbolo

Análise:

Segundo os valores observados podemos verificar que imagens como o 'kid.bmp' apresenta valores bem mais altos de entropia que as restantes imagens. Isto deve-se ao facto de esta imagem apresentar um conjunto mais alargado de símbolos no nosso dicionário (gama de cinzentos) comparativamente às restantes (preto e branco).

Assim sendo estes resultados são esperados porque será preciso considerar mais informação quanto maior a incerteza das ocorrências.

Nos restantes casos o valor da entropia desce pois a incerteza é muito menor, passando apenas por dois valores distintos no caso da 'homerbin.bmp', preto e branco.

No caso da 'guitarSolo.wav' é fácil deduzirmos que a sua entropia seria bastante maior que os exemplos das imagens pois esta apresenta uma vasta gama de ocorrências distintas o que implica uma taxa de compressão mais baixa.

Através destes resultados podemos também concluir que estas imagens podiam ser comprimidas mais eficientemente.

Cada imagem é comprimida com 8 bits/símbolo e a entropia mostramos que é possível comprimir com menos.

No caso da 'homerBin.bmp' o ganho é muito alto, $8 - 0.6448 = 7.3552$. O que significa uma taxa de compressão de 91.94%.

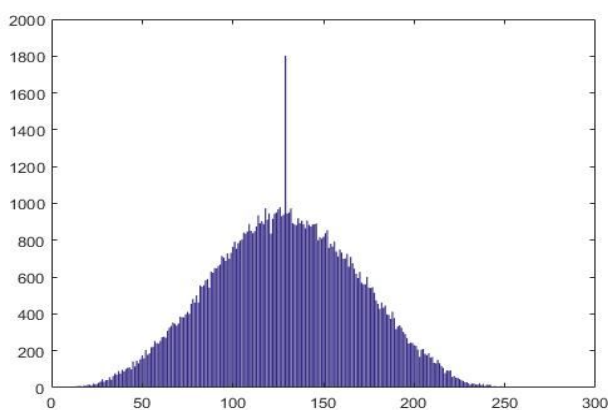


Imagem 1 : Histograma do ficheiro "guitarSolo.wav", sem agrupamento de elementos

Exercício 4:

Resultados obtidos:

kid.bmp: Huffman = 6.9832 bits/símbolo
homer.bmp: Huffman = 3.5483 bits/símbolo
homerBin.bmp: Huffman = 1.0 bits/símbolo
guitarSolo.wav: Huffman = 7.3790 bits/símbolo
english.txt: Huffman = 4.2518 bits/símbolo

Análise:

Como já foi explicado anteriormente o kid.bmp e o guitarSolo.wav possuem uma vasta gama de símbolos que podem ocorrer, a consequência disso será que o Huffman seja mais elevada.

Na imagem homer.bmp existem vários tons de cinzento, no entanto existe um número maior de pixels pretos, isto irá resultar num menor valor de Huffman que irá se aproximar do valor de Entropia.

Contrariamente ao homer.bmp o homerBin.bmp só possui pixels brancos e pretos, ou seja a probabilidade de algum ocorrer é $\frac{1}{2}$ para qualquer um dos símbolos, e aplicando a fórmula $\log_2(2) = 1$.

Partindo do princípio que os valores da entropia são ótimos a variância entre os mesmos e os códigos de huffman deve-se ao facto de que Huffman não gera códigos ótimos, isto é, os valores teóricos são dificilmente alcançados porque é necessário haver todo um conjunto de fatores que estejam de acordo com o esperado. Por exemplo, basta que os símbolos sejam dependentes (o que na maioria é verdade) para que não seja possível chegar a codificação optima.

Esta variância pode ser diminuída fazendo agrupamento de simbolos mas em apenas alguns casos. Ao agruparmos simbolos com menos probabilidade de ocorrência vamos baixar o número de simbolos que temos de codificar, o que resulta variância mais baixa.

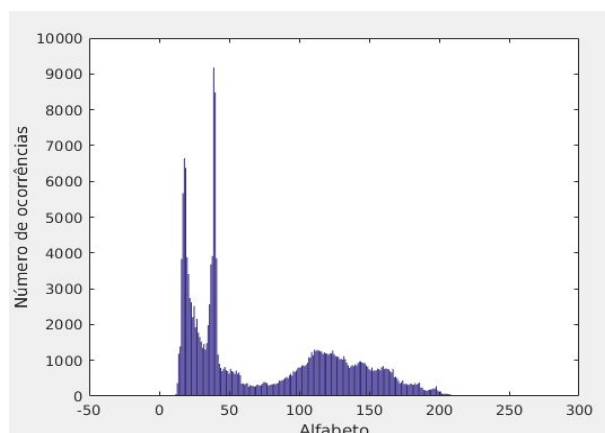


Imagem 2 : Histograma do ficheiro “kid.bmp”, sem agrupamento de elementos

Exercício 5:

Resultados Obtidos:

kid.bmp : Entropia Agrupada: 4.9091 bits/símbolo

homer.bmp : Entropia Agrupada : 2.4127 bits/símbolo

homerbin.bmp : Entropia Agrupada : 0.3978 bits/símbolo

guitarSolo.wav : Entropia Agrupada : 5.7808 bits/símbolo

english.txt : Entropia Agrupada : 3.7588 bits/símbolo

Análise:

Comparando os resultados obtidos no exercício 5 com os do exercício 3, reparamos que a entropia é menor no problema 5, isto acontece pois a entropia de símbolos agrupados será menor do que sem agrupamento, por exemplo, se agruparmos “n” símbolos 4 a 4 ficamos com n^4 resultados possíveis de agrupamentos, fazendo com que seja necessário um maior número de bits para ser codificado. Ou seja, os resultados obtidos são menores pois a compressão de dois números é mais eficaz.

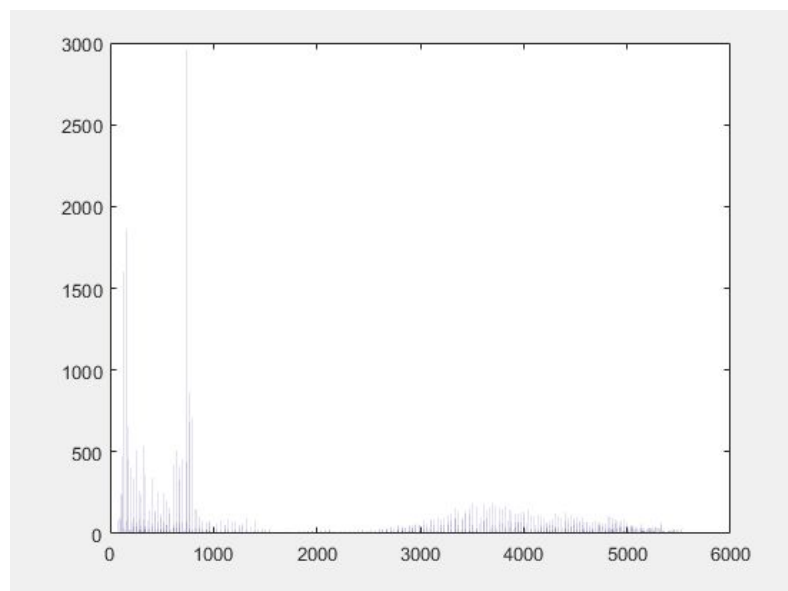


Imagem 3 : Histograma do ficheiro “kid.bmp”, com agrupamento 2 a 2 de elementos

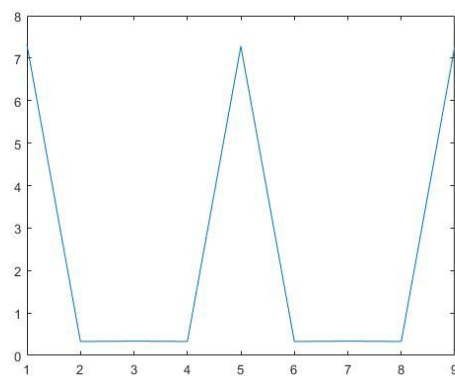
Exercício 6:

Análise:

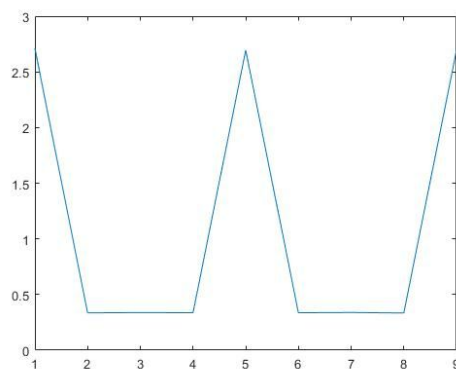
Através da informação mútua é possível determinarmos qual a informação comum em duas fontes de informação, como sinais sonoros. Desta forma com este exercício conseguimos descobrir se um dado sinal conhecido ('guitarsolo.wav') está presente ou se existe alguma semelhança com este ao longo de uma outra faixa sonora.

Após obtermos os resultados reparamos que na faixa 'target01 - repeat.wav' há muito mais semelhança do que na faixa 'target02 - repeatNoise.wav' devido ao facto de que a segunda contém ruído.

Através dos gráficos podemos concluir que apenas partes da faixa têm alguma semelhança com a nossa query, sendo que a restante parte não contém quase nenhuma informação mútua. Esta análise é válida para as duas faixas mas na segunda os valores são ainda mais baixos.



faixa 'target01 - repeat.wav'



faixa 'target02 - repeatNoise.wav'