

Relatório do Trabalho Prático de Teoria da Informação

André Santos

João Botelho

Tiago Martins

Faculdade de Ciências e Tecnologia
da Universidade de Coimbra

1 de Novembro de 2017

Introdução

No âmbito da disciplina de Teoria da Informação, da Licenciatura em Engenharia Informática da Universidade de Coimbra foi proposto a execução de trabalho prático para aprofundar os conhecimentos lecionados nas aulas Teóricas.

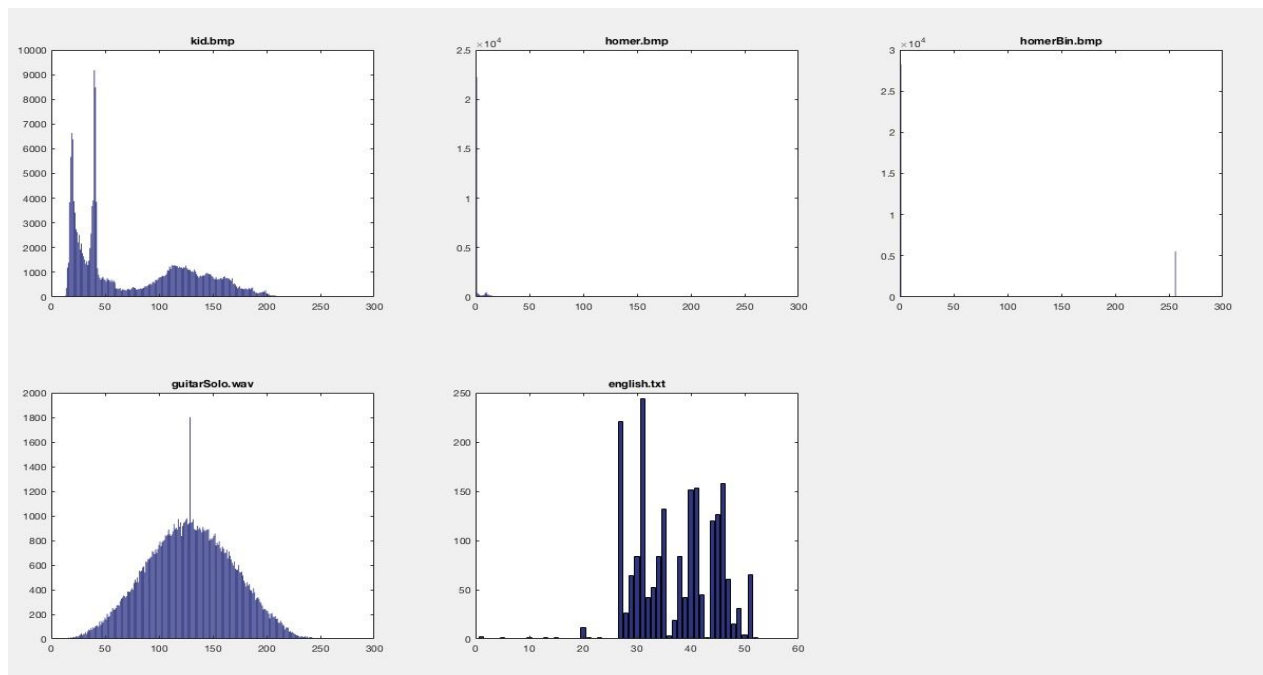
A integralidade do projeto foi na no ambiente de Matlab, que proporcionou a aprendizagem de uma nova linguagem. No trabalho foram implementados 3 conceitos, a entropia, códigos de huffman e a informação mútua.

Distribuição estatística e Entropia

Resultados obtidos para a entropia:

Fonte de Informação	Entropia(bits/símbolo)
kid.bmp	6.954143
homer.bmp	3.465865
homerbin.bmp	0.644781
guitarSolo.wav	7.358020
english.txt	4.228071

Resultados obtidos para a distribuição estatística:



A entropia é o número mínimo de bits necessários para codificar uma fonte de informação. Pelo que é observado nos histogramas, a fonte homerBin tem a menor entropia porque só existem dois valores, 0 e 255. Enquanto a fonte guitarSolo tem a maior devido á utilização de um número maior de símbolos.

Será possível comprimir cada uma das fontes de forma não destrutiva? Se Sim, qual a compressão máxima que se consegue alcançar?

É possível comprimir as fontes de informação de forma não destrutiva em que o limite mínimo de bits por símbolo é o valor da entropia. A taxa de compressão é dada por:

$$\frac{EntropiaMáxima - Entropia}{EntropiaMáxima} * 100$$

Assim podemos calcular o quanto pode ser reduzida.

Código de Huffman

Resultados obtidos para o comprimento médio de bits e a variância:

Fonte de Informação	Comprimento médio de bits	Variância
kid.bmp	6.983223	2.098441
homer.bmp	3.548316	13.196838
homerbin.bmp	1.000000	0.000000
guitarSolo.wav	7.379079	0.756291
english.txt	4.251953	1.109371

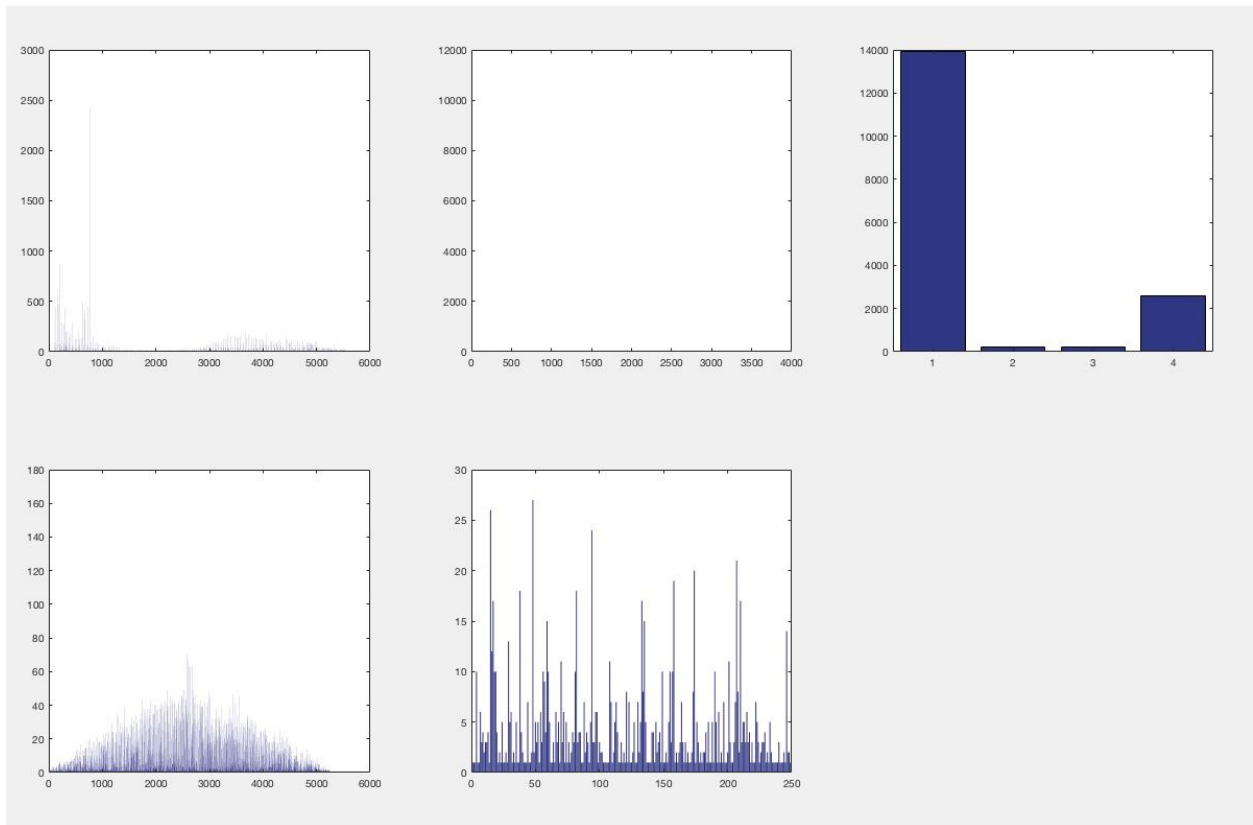
Será possível reduzir-se a variância? Se sim, como pode ser feito em que circunstância será útil?

É possível reduzir a variância, através do agrupamento dos símbolos em pares. Isto é útil para reduzir o número de bits usado para a compressão da fonte de informação.

Entropia para símbolos agrupados

Fonte de Informação	Entropia
kid.img	4.909098
homer.img	2.412733
homerbin.img	0.3978245
guitarSolo.wav	5.7808174
english.txt	3.652150

Frequências dos símbolos agrupados



Através do agrupamento de símbolos foi possível a redução da entropia. O problema é o crescimento exponencial do alfabeto.

Informação Mútua

Resultados obtidos com para a informação mútua:

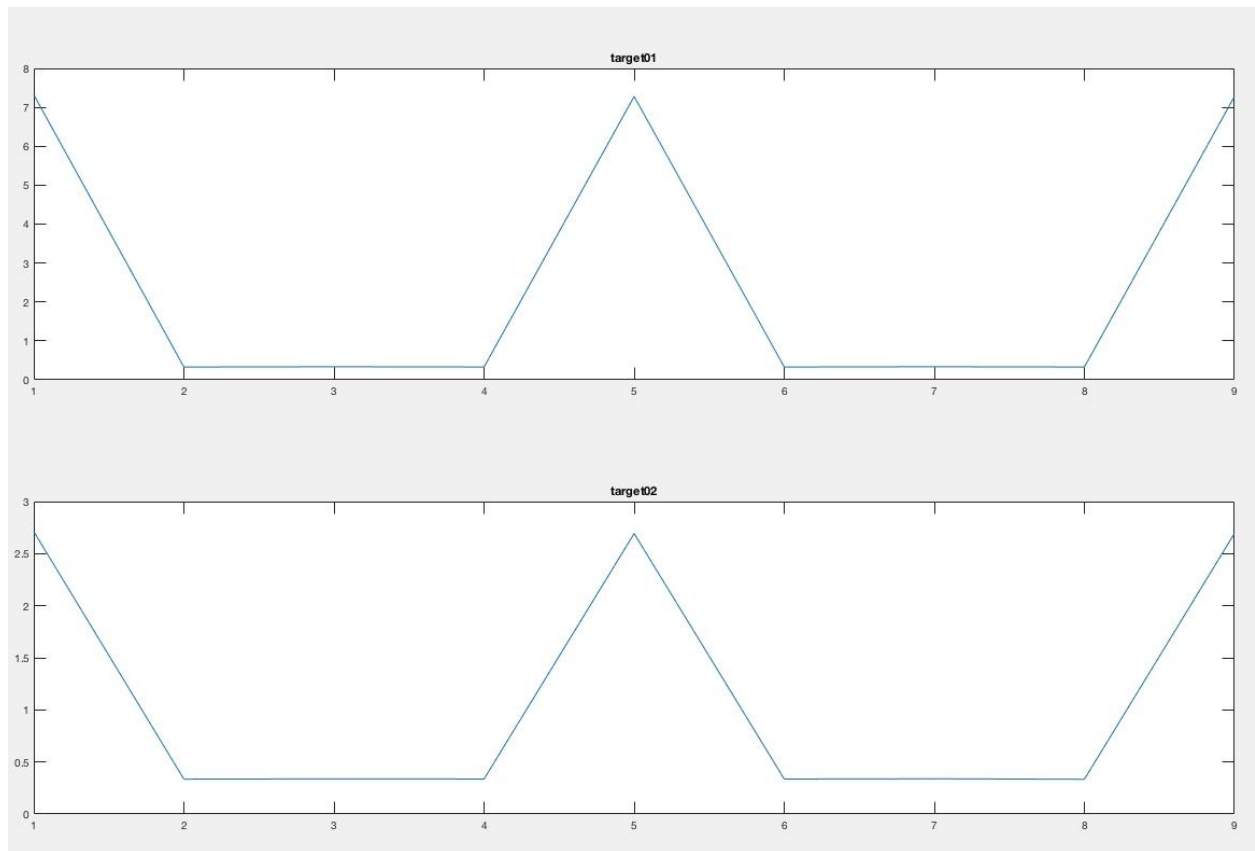
Entre guitarSolo - target01

7.3207 0.3289 0.3354 0.3297 7.2789 0.3289 0.3355 0.3293 7.2806

Entre guitarSolo - target02

2.7113 0.3354 0.3379 0.3366 2.6934 0.3363 0.3388 0.3340 2.6946

Evolução da informação mútua ao longo do tempo:



Devido ao audio ser repetido ao longo do tempo, então isso vai-se refletir no cálculo da informação mútua, por isso vai haver uma espécie de simetria. Informação mútua máxima por ordem decrescente em que guitarSolo.wav é a query e cada uma das fontes do lado esquerdo da tabela são os targets:

Segunda Fonte de Informação	Informação Mútua
Song6	7.3384
Song7	6.3131
Song5	3.9618
Song4	0.4097
Song2	0.3777
Song3	0.3045
Song1	0.2578

A informação mútua aumenta conforme a similaridade das duas fontes. Logo a fonte guitarSolo é mais parecida com a Song6 e menos com a Song1.

Conclusão

A realização deste projeto, propôs um desafio para cada elemento do grupo de trabalho, sendo assim enriquecedor para a aprendizagem destes conceitos chave para a compressão de dados.

Terminando este projeto, alcançou-se o objetivo final. No decorrer do mesmo surgiram algumas dificuldades, sendo a maior calcular a entropia agrupada dos símbolos. Mas com o auxílio dos professores foi possível ultrapassá-las.

Sugerimos assim, a continuidade da realização destes projetos, como forma a desenvolver espírito criativo, indispensável à prática da nossa profissão.