1

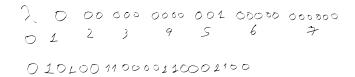
Teoria da Informação - Homework 05

Yuri Niitsuma

EXERCISES.

- 1) The following exercises regard stream codes.

Resposta



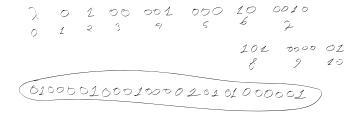
b) (MacKay 6.6) [Medium] Decode the string

00101011110110010010001101010101000011

that was encoded using the basic Lempel-Ziv algorithm.

Resposta

$\underline{001010}\underline{1110110010010001101010101000011}$



2) (The entropy of a compressed file) This exercise regards compression algorithms in general.

An information theory student wants to check whether she can beat Shannon's compression limit of H(X) bits per symbol for an optimal code C applied to a source X.

She envisions a lossless compression method in two steps as follows:

Step 1: Apply an optimal lossless code C to the source X, obtaining a compressed file Y.

Step 2: Apply a new optimal lossless code C' to the source Y, obtaining a new compressed file Z.

Recalling Shannon's Source Coding Theorem, the student makes the following claims about her newly proposed compressing method:

Claim 1: Since code C is optimal for the source X, file Y uses approximately H(X) bits to represent each symbol of X

Claim 2: Since code C' is optimal for the source Y, file Z uses approximately H(Y) bits to represent each symbol of Y

Claim 3: File Z represents each symbol of X using approximately H(X)H(Y) bits.

- a) [Easy] Discuss whether or not each of the student's three claims are correct.
- b) [Medium] What can we say about the size of file Y in comparison to the size of file Z? Is Z gonna be smaller, larger, or of equal size to Y? (Hint: Recall that Shannon's Source Coding Theorem must be valid for the compression from X to Z.)
- c) [Medium] Using your answers to the previous items, what would be an accurate estimation for the value of H(Y)?
- d) [Medium] Using your answers to the previous items, what can the student conclude about the frequency of bits 0 and 1 in any optimally compressed file? How does that relate to the title of this assignment: "Compression and redundancy"?

Resposta

- a) Claims 1 e 2 são colorários do Teorema de Codificação da Fonte.
 Para provar o 3:
 - Cada símbolo de X é representado em Y utilizando H(X) bits.
 - Cada símbolo de Y é representado em Z utilizando H(Y) bits.
 - Cada símbolo de X é representado em Z utilizando H(X)H(Y) bits.
- b) O Teorema de Codificação da Fonte diz que tem que valer a compressão de X pra Z, não importando se o roda em uma, duas ou qualquer quantidade de vezes. Desde, Z não pode ser menor que Y, por outro lado podemos ter uma compressão que bate o limite de Shannon um arquivo utilizando menos que o uso de bits esperado para X por símbolo da fonte. Logo, o tamanho de Z tem que ser aproximadamente o mesmo tamanho de Y.
- c) Do claim 3, Z utiliza H(X)H(Y) bits por símbolo de X e H(X) por símbolo de X. Logo, temos que H(Y) é muito próximo de 1 bit.
- d) $H(Y) \approx 1$ implica que os bits 0 e 1 ocorrem com aproximadamente com mesma frequência em Y. A compressão ótima remove toda redundância da fonte produzindo uma compressão em que a frequência de 0's e 1's é uniforme.