

Teoria da Informação - Homework 04

Yuri Niitsuma

Ainda não sei a tradução correta para *ensemble* e outras, por isso vou deixar o termo original indicado em itálico.

REVIEW QUESTIONS.

- 1) Answer formally the following questions:
 - a) What is a symbol code for an ensemble? What is an extended code for an ensemble?
 - b) When is a symbol code uniquely decodable? When is a symbol code prefix-free?
 - c) State Kraft's inequality and explain in what sense it is related to the notion of which prefix-free codes are actually possible.
 - d) Explain what the Source coding theorem for symbol codes means in terms of the limits of compression of an ensemble.

Resposta

- a) Uma codificação para um *ensemble* X é uma função $C : A_X \rightarrow D^+$ que mapeia cada símbolo $a_i \in A_X$ de um *ensemble* para uma string na codificação do alfabeto D .
- b) Um código de símbolo é *uniquely decodable* se não tem como duas strings formadas por símbolos de um *ensemble* em uma mesa codificação. Um código C é *uniquely decodable* se $\forall x, y \in A_X^+$ temos que $x \neq y \Rightarrow c^+(x) \neq c^+(y)$. Um código de símbolo é livre de prefixo se nenhuma codificação é um prefixo de qualquer *codeword*.
- c) A inequação de Kraft garante que para qualquer código livre de prefixo $C(X)$ sobre um alfabeto tendo D símbolos, o tamanho do código tem que satisfazer a inequação

$$\sum_{i=1}^I D^{-l_i} \leq 1 \text{ onde } I = |A_X|.$$
 Se uma escolha do tamanho da codificação satisfazer a inequação, é possível criar um prefixo com tamanho da palavra-chave escolhida.
- d) O Teorema de Codificação da Fonte (by Wikipédia) para códigos de símbolos indica que para um *ensemble* X , existe um prefixo C com o tamanho esperado limitado inferiormente e superiormente pela entropia $H(X)$

$$H(X) \leq L(C, X) < H(X) + 1$$
 Ou seja, existe uma codificação boa o suficiente que satisfaça a inequação acima.

EXERCISES.

- 2) (MacKay 5.19) [Easy] Is the code $\{00, 11, 0101, 111, 1010, 100100, 0110\}$ uniquely decodeable?

Resposta

Não. A string 111111 pode ser decodificada em três 11 ou duas 111 consecutivas.

- 3) (MacKay 5.20) [Easy] Is the ternary code $\{00, 012, 0110, 0112, 100, 201, 212, 22\}$ uniquely decodeable?

Resposta

O código ternário é livre de prefixo. Logo é *uniquely decodable*.

- 4) (MacKay 5.21) [Medium] Make Huffman codes for X^2 , X^3 and X^4 where $A_X = \{0, 1\}$ and $P_X = \{0.9, 0.1\}$. Compute their expected lengths and compare them with the entropies $H(X^2)$, $H(X^3)$ and $H(X^4)$. Repeat this exercise for X^2 and X^4 where $P_X = \{0.6, 0.4\}$.

You only need to do:

- a) X^2 and X^3 when $P_X = 0.9, 0.1$; and
- b) X^2 when $P_X = 0.6, 0.4$.

Resposta

Defino que o tamanho esperado da codificação C :

$$L(C, X) = \sum_{x \in A_X} p(x)l(x) = \sum_{i=1} p_i l_i$$

- A codificação Huffman para X^2 onde $A_X = \{0, 1\}$ e $P_X = \{0.9, 0.1\}$ é $\{00, 01, 10, 11\} \rightarrow \{1, 01, 000, 001\}$. Este código tem $L(C, X^2) = 1.29$, onde $H(X^2)$ é 0.938.
- A codificação Huffman para X^3 é

$$\{000, 100, 010, 001, 101, 011, 110, 111\} \rightarrow \{1, 011, 010, 001, 00000, 00001, 00010, 00011\}$$

Tendo tamanho esperado $L(C, X^4) = 1.9702$ com $H(X^4) = 1.876$.

- Agora com $P_X = \{0.6, 0.4\}$ os tamanhos para o código de Huffman de X^2 tem tamanho $\{2, 2, 2, 2\}$ e o tamanho esperado é 2 bits (se parar pra pensar se aproxima de $P_X = \{0.5, 0.5\}$), e a entropia é 1.94 bits.
- Para X^4 com $P_X = \{0.6, 0.4\}$:

a_i	p_i	l_i	$c(a_i)$
0000	0.1296	3	000
0001	0.0864	4	0100
0010	0.0864	4	0110
0100	0.0864	4	0111
1000	0.0864	3	100
1100	0.0576	4	1010
1010	0.0576	4	1100
1001	0.0576	4	1101
0110	0.0576	4	1110
0101	0.0576	4	1111
0011	0.0576	4	0010
1110	0.0384	5	00110
1101	0.0384	5	01010
1011	0.0384	5	01011
0111	0.0384	4	1011
1111	0.0256	5	00111

- 5) (MacKay 5.22) [Medium] Find a probability distribution $\{p_1, p_2, p_3, p_4\}$ such that there are *two* optimal codes that assign different lengths $\{l_i\}$ to the four symbols.

Resposta

O conjunto de probabilidades

$$\{p_1, p_2, p_3, p_4\} = \left\{ \frac{1}{6}, \frac{1}{5}, \frac{1}{5}, \frac{2}{5} \right\}$$

nos dá dois diferentes conjuntos de $L(C, X)$ ótimo, porque no segundo passo da codificação de Huffman na árvore de probabilidade qualquer dos três agrupamentos

$$\left\{ \left\{ \frac{1}{6} + \frac{1}{6} \right\}, \left\{ \frac{1}{3} \right\}, \left\{ \frac{1}{3} \right\} \right\}$$

Os dois são:

$$\{00, 01, 10, 11\}$$

ou

$$\{000, 001, 01, 1\}$$

e ambos possuem tamanho esperado 2.

- 6) (MacKay 5.24) [Easy] Write a short essay discussing how to play the game of twenty questions optimally. [In twenty questions, one player thinks of an object, and the other player has to guess the object using as few binary questions as possible, preferably fewer than twenty.]

Resposta

Temos que criar uma classificação em distribua aproximadamente em 50% cada conjunto. De cada classificação, fazemos uma busca binária (de perguntas) em que a árvore criada não pode passar de profundidade 20.