

UNIVERSIDADE DE ÉVORA



ENGENHARIA INFORMÁTICA

TEORIA DA INFORMAÇÃO

Trabalho Prático

Discente:

Marcelo BÁBAU - 30372

José MEDEIRO - 31174

Docente:

Luis RATO

9 de Janeiro de 2016

Conteúdo

1	Introdução	2
2	Desenvolvimento do Programa	3
2.1	Modelo probabilístico da fonte	3
2.1.1	Cálculo das probabilidades	3
2.1.2	Cálculo da Entropia	3
2.1.3	Compressão de dados Shannon-Fano-Elias	4
2.2	Modelo probabilístico do canal	5
2.3	Codificador	6
2.4	Decodificador	6
3	Conclusão	7

1 Introdução

No âmbito da disciplina de Teoria da Informação foi proposto o desenvolvimento de uma aplicação de técnicas de compressão a uma sequência de ADN (sequências de pares base), que tem um alfabeto com 4 símbolos: "A", "C", "T", "G". Para tal implementámos um sistema para a transmissão de informação de uma fonte para um receptor através de um canal ruidoso.

Iremos utilizar um codificador e um decodificador, ligados à fonte e ao canal. O codificador vai consistir em compressão e codificação de dados. Na compressão de dados calculam-se as probabilidades de cada caractere, utilizando a compressão com Shannon-Fano-Elias.

Na decodificação fazemos o oposto da codificação e corrigimos os erros, se existirem.

2 Desenvolvimento do Programa

Durante a realização deste trabalho, foi necessário modelar a fonte: Alfabeto, suas probabilidades, cadeia de Markov. Foi também necessário, modelar o canal: probabilidades de transição.

2.1 Modelo probabilístico da fonte

2.1.1 Cálculo das probabilidades

Alfabeto: "A", "C", "G", "T".

$$P(x = A) = \frac{2374}{11008} = 0.2157$$

$$P(x = C) = \frac{2874}{11008} = 0.2611$$

$$P(x = G) = \frac{3088}{11008} = 0.2805$$

$$P(x = T) = \frac{2672}{11008} = 0.2427$$

2.1.2 Cálculo da Entropia

Entropia da fonte:

$$H(x) = - \sum p(x) \log(p(x))$$

$$H(x) = -[(0.2157 \log(0.2157)) + (0.2611 \log(0.2611)) + (0.2805 \log(0.2805)) + (0.2427 \log(0.2427))] = 1.9934 \text{ bits}$$

2.1.3 Compressão de dados Shannon-Fano-Elias

x	p(x)	F(x)	F'(x)	l(x)	C(x)
A	0.2157	0.2157	0.1079	4	0001
C	0.2611	0.4768	0.3463	3	010
G	0.2805	0.7573	0.6072	3	100
T	0.2427	1	0.8688	4	1110

Figura 1: Compressão de dados Shannon-Fano-Elias

2.2 Modelo probabilístico do canal

O canal que nos foi proposto, é um Canal em Z, ou seja um canal assimétrico.

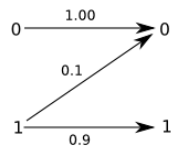


Figura 2: Canal Assimétrico

2.3 Codificador

Para criar o codificador(`Code.java`) tivemos de inicializar dois objetos fundamentais para a codificação. As classes `ShannonFanoElias.java` e `Hamming.java`, são utilizadas respetivamente para nos gerar um código Shannon-Fano-Elias para comprimir a mensagem e depois utilizamos a codificação de Hamming(7,4) para conseguirmos detectar os erros na decodificação.

2.4 Descodificador

O descodificador(`Decode.java`) é o processo inverso do codificador. Depois de recebermos o input, que é a mensagem codificada, voltamos a usar classes `ShannonFanoElias.java` e `Hamming.java`, mas por ordem contrária e a função de descodificar. Em primeiro lugar, detectam-se os erros utilizando a classe `Hamming.java` e retiram-se os bits de paridade. De seguida, utilizando a classe `ShannonFanoElias.java`, fazemos a conversão dos códigos(0001,010,100,1110) de volta para os símbolos("A","C","G","T").

3 Conclusão

Neste trabalho, achamos que não tivemos a totalidade dos objetivos cumpridos, pois tivemos alguns problemas no canal, e o nosso ficheiro com a codificação com os caracteres "0" e "1", não foi gerado como um ficheiro binário, o que nos leva a pensar que o número de erros seja maior.