

Universidad Nacional del Centro de la Provincia de Buenos Aires

Facultad de Ciencias Exactas

Teoría de la Información

Entrega 1: Codificación de la fuente.

Alumnos:

Lacoste, Yamil(libreta 247908).

Luti, Alberto(libreta 247991).

Fecha de entrega: 21 de Abril del 2016.

a)

Para poder obtener la codificación de Huffman para la emisión de una fuente sin memoria, lo primero que se hizo fue leer la imagen y obtener un vector de probabilidades, el cual indica para cada tono obtenido la probabilidad de ocurrencia correspondiente.

Esa probabilidad de ocurrencias se obtiene de la suma de ocurrencia del símbolo dividido el total de elementos que tiene la matriz.

Una vez obtenido el vector de probabilidades, lo que se hizo fue armar el árbol de Huffman, para eso un paso previo fue convertir el vector de probabilidades en un vector de nodos, un nodo es aquel que contiene el símbolo, su probabilidad, el código de Huffman, y dos punteros a sus hijos respectivamente.

Luego, el armado del árbol es un procedimiento que va desde el vector de nodos llamado "hojas" hasta la raíz, la idea es ir sacando las menores probabilidades del vector y armar un nuevo nodo con la suma de esas dos probabilidades. Ese nuevo nodo, se le asigna los punteros a cada uno de los nodos sumados, para la obtención del código de Huffman al nodo de menor probabilidad se le pose un cero, y al otro un uno respectivamente.

Una aclaración importante, es que ese uno o cero que se le asigna en el nodo perteneciente al código de Huffman es propagados a cada uno de sus hijos, lo cual una vez armado el árbol de Huffman ya se obtiene en los nodos "hojas" el código de Huffman correspondiente a cada símbolo.

b)

Para una fuente con memoria de orden 2, lo que se hizo fue leer la imagen nuevamente, obtener el vector de probabilidades para un símbolo que es lo que calculamos en a), y también se armó la matriz condicional, ambas mirando los símbolos leídos de la imagen.

Para el armado de la matriz condicional lo que se hace es, calcular las probabilidades de ocurrencia de sacar un símbolo sabiendo que salió tal símbolo, ya que estamos calculando de orden dos.

Para eso, lo que hicimos era tener una matriz y un vector, los cuales íbamos aumentando en uno sus valores, el vector se aumentaba en la referencia del primer símbolo obtenido. Por su parte, la matriz se aumentaba en la posición de los dos símbolos. Al finalizar, se dividió la matriz con el vector y se obtuvo una matriz resultante llamado "condicional".

Por ultimo para tener el vector de probabilidades obtenido en a) lo que se hace es multiplicar el vector de probabilidades de a) (que es la probabilidad de ocurrencia de cada símbolo) por la matriz condicional (la cual es la probabilidad de que sabiendo que salió un símbolo salga tal otro símbolo).

Luego ese vector obtenido, se lo prosigue de igual forma que lo hecho en el inciso a), es decir se arma el árbol y se obtienen los valores de Huffman asociados a cada símbolo correspondiente, los cuales se encuentran en las hojas del árbol.

c)

El cálculo de la longitud media es igual para cada inciso lo que se hizo fue:

$$a) \frac{\langle Lm \rangle}{n} = \frac{\sum_{k=1}^m Lk * Pk}{n}$$

Es decir recorrer el vector de las raíces e ir sumando en una variable final, la longitud de cada símbolo por su probabilidad de ocurrencias.

Para a) como tenemos orden 1, el n=1 y para B) el n=2.

d)

Entropía para a)

$$H * = - \sum_{i=1}^n (\log_2 P_i) * P_i$$

El h* la cual se obtiene de es recorrer el vector de probabilidades y hacer la sumatoria del logaritmo de esa probabilidad por la probabilidad dada.

Entropía para b)

$$H_{cond} = \sum_{j=1}^m P(s_j) \sum_{i=1}^n P(si \setminus sj) * \log_2 P(si \setminus sj)$$

Se calculó multiplicando el vector de probabilidad de obtener el primer símbolo por la matriz condicional por el logaritmo del valor de la matriz condicional.

Entropía condicional del par Que se obtiene restando lo obtenido en a)-b):

$$H1 = H * - \sum_{j=1}^m P(s_j) \sum_{i=1}^n P(si \setminus sj) * \log_2 P(si \setminus sj)$$