Teoría de la información

Trabajo practico especial 2023

**Número de grupo:** 23

**Alumnos y libreta universitaria:**

* Koprivica, Savo.
* Luna, Demetrio. 249856
* Piu, Florencia. 249943

**Fecha:** 22/6

**Introducción**

Se analizarán dos señales de sensores de luminosidad en dos telescopios en distintas ubicaciones del país, estos documentos fueron dados por la catedra. Mediante el análisis de estas señales se pretende obtener información del comportamiento de las mismas. Este estudio se basará en la distribución de probabilidades obtenida en cada señal y su codificación según el método de Huffman.

**Codificación Huffman**

**La entropía**

La entropía es el mínimo número de preguntas binarias que hay que hacer en promedio para identificar un símbolo de la fuente.

Para calcular la entropía en nuestras señales lo primero que se hizo fue obtener la distribución de probabilidad de cada símbolo dividiendo la cantidad de ocurrencia de cada uno entre la cantidad de símbolos totales de la fuente. Para luego multiplicar cada una de estas probabilidades por la longitud optima de cada símbolo, que esta se obtiene realizando la siguiente formula , siendo pk la probabilidad de ocurrencia del símbolo. Por lo tanto, la entropía quedaría definida por la siguiente formula.

* La entropía de la Señal 1: el número mínimo de preguntas binaria en promedio que hay que realizar es de 9.5851 para obtener un símbolo.
* La entropía de la Señal 2: el número mínimo de preguntas binaria en promedio que hay que realizar es de 6.6697 para obtener un símbolo.
* Esta diferencia en la entropía entre las señales nos quiere decir que se necesitan menos preguntas binarias para generar un símbolo en la señal 2.

**La longitud media de la codificación Huffman**

La longitud media de la codificación Huffman es la longitud media en bits que se utilizara para cada símbolo. Esto se obtiene multiplicando la probabilidad de ocurrencia de cada símbolo por la longitud en bits de cada símbolo.

* Longitud media de señal 1: La longitud media en bits que se utilizara para cada símbolo es de 9.6320.
* Longitud media de señal 2: La longitud media en bits que se utilizara para cada símbolo es de 7.059.
* En promedio se necesita aproximadamente 2.53 bits más para codificar un símbolo de la señal 1 que de la señal 2, por esto podemos entender que en la mayoría de los casos la señal 1 va hacer más grande que la señal 2 en tamaño.

**Diferencia de bits entre la codificación Huffman y el archivo original**

Para realizar esta comparación se toma en cuenta el peso original del archivo en bits y el peso del mismo archivo comprimido mediante el método Huffman. Para obtener el tamaño de este último se calcula la codificación Huffman de cada símbolo y luego se transforman las señales originales a este código y se cuentan la cantidad de bits que hay.

* La cantidad de bits que se necesita para la codificación Huffman de la señal 1 es de 9632 bits y el tamaño original es de 55984 bits. La tasa de compresión es de 1:5,81.
* La cantidad de bits que se necesita para la codificación Huffman de la señal 2 es de 7059 bits y el tamaño original es de 1268600 bits. La tasa de compresión es de 1:179,71.

**Rendimiento del código Huffman y la demostración de Shannon**

El rendimiento del código Huffman se refiere a la eficiencia de compresión que se logra al utilizar este algoritmo. El rendimiento se calcula mediante la siguiente formula . Siendo H la entropía y L la longitud media del código.

El primer teorema de Shannon estipula que la longitud media del código de Huffman va hacer menor o igual a la entropía de la fuente y menor estricto que la entropía de la fuente más uno.

El teorema de Shannon para fuentes de orden extendida estipula que la longitud media del código de Huffman cuando el orden tiende a infinito, la diferencia entre la entropía y la longitud media del código Huffman termina siendo despreciable. Por lo tanto, la longitud media del código Huffman tiende a hacer igual a la entropía.

* La eficiencia de compresión de la señal 1 es de 0,995.
* La eficiencia de compresión de la señal 2 es de 0,945.
* Tiene mejor rendimiento la señal 1,
* Demostración de Shannon de la señal 1:

9.5851 ≤ 9.6320 < 9.5851 + 1

* Demostración de Shannon de la señal 2:

6.6697 ≤ 7.059 < 7.059 +1

**Valor de media y desvió estándar**

El valor de media, también conocida como promedio, es una medida estadística que representa el valor central de un conjunto de datos. Se calcula sumando todos los valores y dividiendo el resultado entre el número total de valores.

El desvío estándar, también conocido como desviación estándar, es una medida estadística que indica la dispersión o variabilidad de un conjunto de datos con respecto a su media. Se utiliza para determinar cuánto se alejan los valores individuales del promedio.

* El valor de la media(promedio) de la señal1 es 32564.87 y el desvió estándar es de 15605.22 lo cual nos indica que los valores de la señal 1 están muy dispersos respecto de su media.
* El valor de la media(promedio) de la señar2 es 1066.1 y el desvió estándar es de 554.78 lo cual nos indica que los valores de la señal 2 están muy dispersos respecto de su media

**Conclusión**