

Universidad Politécnica de Cartagena



Escuela Técnica Superior de Ingeniería de Telecomunicación

PRÁCTICAS DE TRANSMISIÓN DE DATOS

Práctica 1: Entropía de una fuente de información

Profesor:

Javier Vales Alonso

1. INTRODUCCIÓN

En esta práctica se implementará en matlab un **sistema para el cálculo de la entropía** de una fuente dada (nosotros usaremos como fuente ficheros de datos). El objetivo es profundizar en el estudio y la comprensión del concepto de la **información** como magnitud medible y presentar las bases para el desarrollo de futuras prácticas.

2. OBJETIVOS

Teóricos:

- Asimilar el concepto de entropía como verdadero representante de la información obtenible de una fuente.
- Discernir entre tasa y información.

Prácticos:

- Obtener a partir de una fuente dada (p. ej. variables aleatorias, un fichero de texto, un fichero de imagen, un *pipe* Unix) la entropía por carácter.
- Determinar la entropía de un texto en castellano a partir del Quijote y compararla con otras fuentes.

3. DESARROLLO DE LA PRÁCTICA

La práctica se desarrollará en una única sesión y se divide en las siguientes partes:

1. Primero se implementará una función de matlab (**entropia.m**) que permita calcular la entropía de una variable aleatoria a partir de las frecuencias relativas de la aparición de símbolos.
2. A continuación se implementará otra función (**calculo de frecuencias.m**) que a partir de una fuente dada obtenga la frecuencia relativa de los símbolos.
3. Tras ello se aplicará a varios ficheros de ejemplo los procedimientos anteriores y se analizarán los resultados obtenidos.
4. Finalmente, se propondrá un ejercicio de ampliación.

Para la realización de esta práctica deberán descargarse en la carpeta de trabajo los ficheros de ejemplo disponibles en el apartado correspondiente del aula virtual.

3.1. Implementación de la función *entropia*

Como hemos estudiado en teoría, el uso de la entropía como medida de la información de una variable aleatoria surgió de los estudios de Claude Shannon para intentar descubrir cuál era la mayor tasa de transmisión de datos alcanzable sin errores. La definición de entropía (que guarda relación con la conocida magnitud termodinámica) es:

$$H = - \sum_{i=1}^N p_i \log_2(p_i)$$

Siendo las p_i las probabilidades de que el símbolo tome el valor i -ésimo (es decir la masa de probabilidad de una variable aleatoria X).

EJERCICIO: Debe crear una función *entropia* que realice el cálculo anterior

```
function H = entropia(freq)
% Calcula la entropia de una variable aleatoria X cuya función
% de masa esta dada por el vector freq

end
```

(LOS CODIGOS DEBE ENTREGARLOS EN LA MEMORIA)

Verificación: (A ENTREGAR EN LA MEMORIA LAS RESPUESTAS Y EL CÓDIGO QUE EJECUTE PARA LLEGAR A ELLAS)

- Consideremos como variable X el lanzamiento de una moneda, ¿cuál es la entropía de una tirada?
- ¿Y para un dado de 6 caras?
- ¿Cuántas caras debería tener un dado para tener una entropía de 6 bits? Justifique la relación del resultado con la cantidad de permutaciones obtenibles con 6 dígitos binarios.
- ¿Y para que la entropía sea de al menos 5 nats?
- ¿Qué sucede cuándo la probabilidad de un valor concreto, p_i , es nula? Justifíquelo analíticamente.
- Considere en los ejemplos anteriores un desplazamiento de la función de, es decir, se trata de medir la entropía de Y , tal que $p(Y=i) = p(X=i-D)$, con D fijo. ¿Qué sucede con la entropía? ¿por qué?
- ¿Y en caso de invertir el orden, $p(Y=i) = p(X=-i)$?
- ¿Qué puede decir en general sobre una permutación cualquiera de la masa?
- Sea X una variable aleatoria binomial $B(30, p)$. Es decir, su masa es:

$$p_i = \binom{n}{i} p^i (1-p)^{(n-i)}$$

para $n=30$. Calcule su entropía para p en el rango 0:0.1:1, represéntela en una gráfica y explique el resultado.

3.2. Frecuencia relativa de los símbolos de una fuente

En muchas situaciones prácticas interesa determinar la entropía de una fuente de la que no se tiene una caracterización estadística analítica (como en las cuestiones anteriores), sino que hay de obtenerse la masa a través del calculo de la frecuencia relativa de los símbolos de una fuente. En este apartado trabajaremos usando como fuente archivos de datos, y cómo símbolo trabajaremos ahora mismo a nivel de carácter (grupo de 8 dígitos binarios).

EJERCICIO: Debe crear una función *calculofrecuencias* que calcule la frecuencia de aparición de los caracteres en un fichero dado

```
function freq = calculofrecuencias ( nombre_fichero )
% Calculo de la frecuencia de aparicion de cada caracter
% en el fichero indicado. Devuelve array de 256 posiciones
% indexado por el caracter correspondiente

. . . % codigo

% Incluir representacion grafica de las frecuencias
figure;
plot(freq);

end
```

(LOS CODIGOS DEBE ENTREGARLOS EN LA MEMORIA)

3.3. Pruebas y análisis de los resultados

(A ENTREGAR EN LA MEMORIA LAS RESPUESTAS Y EL CÓDIGO QUE EJECUTE PARA LLEGAR A ELLAS)

- Justifique y adjunte a la memoria las gráficas de las frecuencias relativas de cada uno de los ficheros.
- ¿Cuál es la entropía de cada uno de los ficheros de ejemplos?
- ¿Por qué la diferencia entre dos formatos que contienen la misma foto?
- A la vista de los resultados, ¿cuál es el límite mínimo de tamaño (en dígitos binarios) alcanzable mediante compresión de cada uno de los ficheros? Justifique la respuesta.

3.4. Ejercicio: Símbolos extendidos

Por último, proponemos una variante sobre el método anterior. Usaremos como concepto de símbolo una agrupación de K caracteres y no un solo carácter.

EJERCICIO: Debe crear una función *calculofrecuenciasbloque* que calcule la frecuencia de aparición de los bloques de caracteres en un fichero dado

```
function freq = calculofrecuenciasbloque ( nombre_fichero, bloque )
% Calculo de la frecuencia de aparicion de cada caracter
% en el fichero indicado. Devuelve array de 2^(8*bloque) posiciones
% indexado por el ordinal del bloque correspondiente

. . . % codigo

% Incluir representacion grafica de las frecuencias
figure;
plot(freq);

end
```

(LOS CODIGOS DEBE ENTREGARLOS EN LA MEMORIA)

(A ENTREGAR EN LA MEMORIA LAS RESPUESTAS Y EL CÓDIGO QUE EJECUTE PARA LLEGAR A ELLAS)

Verificación:

- Repita los ejercicios de 3.3. ¿A qué se deben las diferencias observadas?
- Proporcione una gráfica para el Quijote de la entropía para bloques de tamaños 1 a 4 caracteres. ¿Qué interpretación da a que la gráfica sea decreciente?