Introducción a la Teoría de la Información

Resumen de definiciones + Ejemplos

Alicia Fernandez

A pesar de que la teoría de señales es una herramienta valiosa, no se centra en el proceso fundamental de las comunicaciones que es la transferencia de información. En 1948 Claude E. Shannon desarrolló una teoría[1] que después dió lugar a una nueva disciplina: *Teoría de la Información*.
El cuestionamiento que se planteó Shannon es, *"...dada una fuente productora de mensajes que no son de nuestra elección, cómo se deben representar los mensajes para obtener una comunicación confiable sobre un canal con limitaciones físicas."*

Teorema fundamental de Shannon para canales con ruido

Si la cantidad de información de la fuente no excede la capacidad del canal existe una técnica de codificación tal que la información puede ser trasmitida con una probabilidad de ocurrencia de errores arbitrariamente pequeña, a pesar de la presencia de ruido.

Haremos las definiciones para el caso discreto y generalizaremos para el caso continuo, en este repartido no veremos demostraciones.

Medida de la Información

Usaremos el término *información* como un término técnico, el que no debe confundirse con conocimiento o significado.

Definición: La información que envía una fuente digital cuando transmite el mensaje j-ésimo está dada por:

$$I_j = \log_b \frac{1}{P_j}$$

P_j es la probabilidad que el mensaje j-ésimo sea seleccionado por la fuente para ser trasmitido.
Observación: I_j información propia del mensaje x_j, independiente del contenido o la posible interpretación del mensaje. La base del logaritmo detemina las unidades de medida de la información, si $b = e$ la unidad es *nats*, si $b = 2$ la unidad es el *bits*, y si $b = 10$ la unidad es el *hartley*.
Bit aquí se usa como unidad de información y no como unidad de dato binario (a veces se usa *binit* para no confundir).
En el caso de una fuente que emite dos mensajes equiprobables cada dato binario lleva un bit de información. Si $P_1 = 1/4 \Rightarrow I_1 = 2$ mientras que $P_0 = 3/4 \Rightarrow$

[1] *A Mathematical Theory of Communication.* C.E.Shannon. Bell System Technical Journal. Julio 1948

$I_0 = 0.41$. El bit 1 lleva 2 bits de información mientras que el 0 lleva 0.41 (menor, por ser más probable).

Consecuencias de la definición

1. $I_j \geq 0$ si $0 \leq P_j \leq 1$
2. $I_j \rightarrow 0$ si $P_j \rightarrow 1$
3. $I_i < I_j$ si $P_i > P_j$

Si suponemos los mensajes independientes, $P(x_i, x_j) = P_i P_j \Rightarrow$

$$I_{ij} = \log_2 \frac{1}{P_i P_j} = \log_2 \frac{1}{P_i} + \log_2 \frac{1}{P_j} = I_i + I_j$$

Interpretación: Los mensajes menos probables son los que llevan más información.
Por ejemplo si estamos por viajar y queremos saber el estado del tiempo y escuchamos uno de estos tres mensajes: Amanecerá – Lloverá – Tornado
El primer mensaje no agrega información al receptor, pues no existe incertidumbre sobre ese punto. Mientras que el segundo si lleva alguna información (puede ser mucho mayor en Palma que en Londres). El último mensaje que es el más improbable es el que trasmite más información por lo inesperado y nos puede llevar a suspender el viaje.

Si la fuente es de memoria nula, es decir, los mensajes sucesivos son independientes, la información que obtengo al recibir dos mensajes es la suma de informaciones individuales.

Entropía

Asumiremos que la fuente es estacionaria, por lo tanto las probabilidades se mantienen constantes en el tiempo y que símbolos sucesivos son estadísticamente independientes y se trasmiten a una velocidad promedio (cadencia) de r símbolos por segundo. Estas hipótesis corresponden al modelo de fuente discreta sin memoria.

Sea $X = \{x_1, x_2, \ldots, x_N\}$ una fuente que emite un conjunto de N símbolos.

Definición:

$$H(x) = E\{I(x)\} = \sum_{i=1}^{N} P_i I_i = \sum_{i=1}^{N} P_i \log_2 \frac{1}{P_i} \quad \frac{bits}{simbolo}$$

Observar que $0 \leq H(x) \leq \log_2 N$

La entropía sólo depende de las probabilidades y del tamaño del alfabeto.
Analizaremos el caso particular de $N = 2$, para ver como varía $H(X)$. El 0 corresponde a tener $P_1 = 1$ y $P_0 = 0$ (o viceversa) es decir el caso en el que no hay incertidumbre por parte del receptor, mientras que el otro extremos corresponde al caso de símbolos equiprobables que corresponde con la máxima incertidumbre.

Velocidad de transferencia de información de la fuente

Definición:

$$R = rH(x) \ \ bits/seg$$

con $r : simbolos/seg$

Ejemplo. Supóngase que una fuente emite con $r = 2000 \ \ simbolos/seg$ seleccionados de un alfabeto de tamaño $M = 4$ con probabilidades $P_i = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$. Determinar la velocidad de transferencia de información.

La información por símbolo es, $Ii = \{1, 2, 3, 3\}$.
La entropía de la fuente es, $H(X) = \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + \frac{1}{8} \times 3 = 1.75 \ \ bits/simbolos$
Entonces $R = 2000 \times 1.75 = 3500 \ \ bits/seg$ $\diamond$

Vamos a ver que pasa cuando trasmitimos a través de un canal con ruido, seguiremos con las hipótesis de que tanto la fuente como el canal son discretos de modo que podemos medir la cantidad de información transferida y definir la capacidad del canal.

Información mutua

X: alfabeto fuente, la fuente selecciona símbolos de este alfabeto para su trasmisión sobre el canal
Y: alfabeto destino, no tiene porque ser el mismo que el X si en el canal se introduce ruido.
Canal: invariante en el tiempo y sin memoria.
$P(x_i)$: probabilidad de que la fuente seleccione el símbolo x_i para trasmisión
$P(y_j)$: probabilidad de que el símbolo y_j se reciba en el destino
$P(x_i \ y_j)$ probabilidad conjunta de que x_i sea trasmitido e y_j recibido
$P(x_i|y_j)$: probabilidad condicional de que x_i sea trasmitido dado que se recibió y_j
$P(y_j|x_i)$: probabilidad condicional de que y_j sea recibido dado que se trasmitió x_i. Se les llama probabilidades de transición hacia delante del canal.

Definición:

$$I(x_i; y_j) = \log_2 \frac{P(x_i|y_j)}{P(x_i)}$$

Representa la cantidad de información que se transfiere cuando x_i se trasmite e y_j se recibe.

$$I(X; Y) = \sum_{X,Y} P(x_i \ y_j) I(x_i; y_j) = \sum_{X,Y} P(x_i \ y_j) \log_2 \frac{P(x_i|y_j)}{P(x_i)}$$

Representa la cantidad promedio de información proveniente de la fuente obtenida por símbolo recibido.

Interpretación: en un canal ideal sin ruido para cada x_i hay un y_j tal que $P(x_i|y_j) = 1$, entonces la información mutua iguala a la información propia. Mientras que si el canal es muy ruidoso y x_i e y_j no están correlacionados $P(x_i|y_j) = P(xi)$ y $I(x_i; y_j) = 0$, no se transfiere información.

Equivocación o Entropía condicional

Definición:

$$H(X|Y) = \sum_{X,Y} P(x_i\ y_j) \log_2 \frac{1}{P(x_i|y_j)} = H(X) - I(X;Y)$$

Interpretación: representa la información perdida en un canal con ruido. En un canal sin errores $H(X|Y) = 0$. Se puede demostrar que en presencia de ruido $H(X|Y) > 0$

Ejercicio: Sea una secuencia binaria tal que $P_1 = P_0 = \frac{1}{2}$ y el error promedio del canal es $P_b = 0.01$. Hallar la equivocación y la información efectiva trasmitida.

Capacidad de un canal discreto

Definición:

$$C_s = max\{I(X;Y)\}\ \frac{bits}{simbolos}$$

Si s es la velocidad máxima de transferencia permitida por el canal,

$$C = sC_s$$

Donde C representa la capacidad por unidad de tiempo.

Dado un canal, si los alfabetos de entrada y salida están fijos lo mismo que las probabilidades de transición, la forma de maximizar $I(X;Y)$ es variando las $P(x_i)$.

Ley de Hartley-Shannon

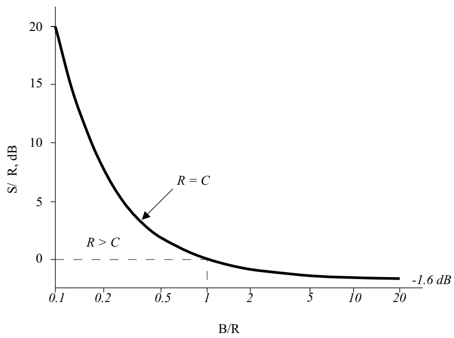
Para canales con ruido aditivo blanco gaussiano:

$$C = B \log_2(1 + S/N) = B \log_2(1 + S/\eta B)$$

La curva de la figura especifica el intercambio óptimo entre ancho de banda.
Definimos un sistema de comunicación ideal como aquel que puede alcanzar una trasmisión libre de errores a una velocidad:

$$R = B \log_2(1 + S/N).$$

Este sistema es físicamente irrealizable.



Referencias

- *Communication Systems*. Bruce A. Carlson. 4th. edition. Capítulo 16 (§16.1 y §16.2)
- *Information Theory and Coding*. N. Abramson. McGrawHill, NY