DOI: http://dx.doi.org/10.30972/eitt.604394

Teoría de la información y codificación: El significado de la entropía en la transmisión de información.

Mag. Ing. Emilio Fabián Scozzina(*)

Resumen.

Desarrollaré algunas de las ideas más interesantes en uso del concepto de entropía, aplicado en la ***teoría de información y codificación.*** Este siempre ha sido un tema discutido y que da muchas interpretaciones del fenómeno, según el observador se encuentre posicionado (desde la fuente de información ù observando la misma.

El objetivo de esta publicación, es presentar algunos aspectos, que de describen matemáticamente el concepto básico de entropía, dado que la entropía cobra singular importancia en el estudio de cualquier fuente de información o la capacidad de los canales, también está relacionada con la incertidumbre que existe en cualquier señal de tipo aleatoria.

Palabras Claves: Entropía, Información, Codificación,

1. Introducción.

Información

La definición de información fue elaborada por Claude E. Shannon, y presentada en su histórico trabajo *"A Mathematical Theory of Communication (1948)"* fue con el que sentó las bases de las matemáticas de la teoría de la información y las comunicaciones modernas, aplicando el álgebra de Boole, este estudio que dio el ***fundamento matemático para industrializar el procesamiento de la información,*** tan vigente hoy en día.

Partimos de la definición de Información, sea un suceso cualquiera [E] el cual puede presentarse con una probabilidad P(E), una vez ocurrido dicho suceso, la información asociada al mismo definimos como I(E). La información es inversamente proporcional a la

(*) Contacto: efscozzina@gmail.com | Tel. Celular: 0362-15-4527366.

probabilidad de ocurrencia de un suceso, veamos aquí la expresión matemática. Abramson (1966).

$$I(E) = log_2 \frac{1}{P(E)} \; expresada \; en \; bits$$

Según sea la base del logaritmo utilizado, esta determinara la unidad de medida de información.

Evolución de las unidades la información.		
nat, nit ò nepit	*ln* (Base es e)	Es la más antigua pero sigue vigente.
Hartly – ban-dit	*log* (Base 10)	Hartely (1928)
bits	log_2 (Base 2)	Shannon (1948)
cubit - qbit	Esfera de Bloch	Schumacher (1995)

Tabla: Unidades de Entropia de Información IEC 80000-13:2008

Tabla: Unidades de Entropía de Información IEC 80000-13:2008

En cuanto a la Entropía, la definición de la función entropía, H(S), es el producto de la probabilidad de ocurrencia de un evento por la información asociada a dicho evento *H(S)=I(E) P(E)*. La entropía, *H(S)*, es una magnitud que mide la información por símbolos emitidos desde una fuente.

Esta definición puede aplicarse a fuentes de información de cualquier naturaleza, y nos permite por ejemplo codificarla o decodificarla adecuadamente, indicándonos los elementos de código necesarios para transmitirla de manera óptima eliminando toda redundancia (código instantáneo), o determinar las equivocaciones en un canal de información.

Una rápida interpretación nos permite decir que la entropía de una fuente, es la ***cantidad de bits por símbolos emitidos en dicha fuente***. Si lo analizamos desde el punto de vista de un observador, que está estudiando dicha fuente, la entropía es la ***incertidumbre*** que tiene el mismo respecto del próximo símbolo que será emitido.

La medida de información debe ser proporcional, un cambio pequeño en una de las probabilidades de aparición de uno de los símbolos de la señal, hace cambiar la entropía.

$$H(S) = \sum_{i=1}^{s} P(E)\, I(E) = \sum_{i=1}^{s} P(E)\, log_2 \frac{1}{P(E)} \; expresada \; en \; \frac{bits}{simbolos}$$

Para ahondar más en el concepto de entropía, consideramos un experimento: como ser un ***"ensayo de Bernoulli"***, es un ejemplo clásico muy utilizado en la teoría de probabilidad y estadística.

Es un experimento de tipo aleatorio, en el que cual solo se pueden se puede obtener dos resultados, como es caso de lanzar una moneda al aire, estos resultados pueden ser cara o cruz.

Lo mismo ocurre con generador binario aleatorio, o un modulador 2-ASK aleatorizado, el cual arroja unos y ceros de manera equiprobables. Aquí la variable aleatoria puede tomar dos valores 0 y 1 lógicos, (cara o cruz), esta es una función que asigna un valor, usualmente numéricos a resultados de un experimento.

Veamos cómo funciona este experimento: Si asignemos valores, supongamos que *p* es la probabilidad de ocurrencia de un 1 lógico, y (1-*p*) = q, 0 lógico, entonces el valor del valor esperado E[X] de la variable aleatoria es *p* y su varianza, *p* (1-*p*).

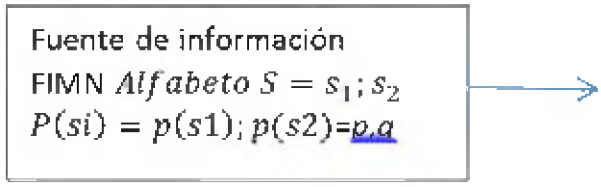
Los procesos de Bernoulli son los que resultan de la repetición en el tiempo de ensayos de Bernoulli independientes pero idénticos.

Parámetros	$0 < p < 1, p \in \mathbb{R}$ $caso\ particular\ p = 0{,}5$
Dominio	$k = [0,1]$
Función probabilidad f(p)	$q = (1-p)\ para\ k = 0\ y\ p\ para\ k = 1$ Caso particular, analizado p=q=0,5
Distribución de probabilidad	"0" $para\ valores\ de\ k < 0$, no está definida para valores negativos. $q\ para\ 0 \leq k < 1$ $p\ para\ k \geq 1$
Varianza	$p \times q = 0{,}5 \times 0{,}5 = 0{,}25$
Moda	0 y 1 si q = p (los dos valores)
Curtosis para p=0,5 la distribución de Bernoulli tiene un valor menor que el de cualquier otra distribución.	$\frac{6p^2 - 6p + 1}{p(1-p)} =$
Caracterización Binomial $X \approx B\ (n,p)$	$P(X = 0) = f(0) = 0{,}5^0 \cdot 0{,}5^1 = 0{,}5$ $P(X = 1) = f(1) = 0{,}5^1 \cdot 0{,}5^0 = 0{,}5$

La información asociada esta fuente será un bit:

$$I(0) = log_2 \frac{1}{P(0)} = I(1) = log_2 \frac{1}{P(1)} = log_2 \frac{1}{0,5} = 1 \ bits$$

Podemos obtener como resultado el bit 0 ó el bit 1. Este experimento puede asimilarse a una (FIMN), *fuente de información de memoria nula*, en cada instante de tiempo **t** genera un símbolo $\mathbf{s_i}$ elegido dentro del conjunto de símbolos posibles -*o alfabeto fuente*- S {0,1}, según sus probabilidades de emisión (modelo p=0,5; q=0,5). La entropía está asociada a cuál es la información promedio que emite una fuente, es de 1 bits/símbolo.



Los símbolos de la fuente son estadísticamente independientes, cumple con los axiomas de Kolmogorov (1933), por ello la fuente se describe mediante el conjunto de símbolos posibles $\mathbf{s_i}$ y la probabilidad de ocurrencia $\mathbf{p\ (s_i)}$ asociada a cada símbolo.

Este modelo nos permite definir, una fuente de información de memoria nula. La probabilidad de emitir un símbolo depende solo del símbolo emitido y no del que fue emitido en el instante anterior. No existen probabilidades condicionales $\mathbf{p\ (s_j\ /s_i) = 0}$ entre los símbolos emitidos. Las probabilidades de emisión son constantes en el tiempo, para este caso.

Esto se podría asimilar a un proceso de marakov con memoria nula, o de orden 0. Los estados de la fuente evolucionan en el tiempo de manera impredecible, desde el punto de vista del observador, ***en este punto la entropía es máxima.***

¿Qué ocurre si las probabilidades varían? Ahora **p** no es igual a **q,** pero siempre se cumple que la suma de **(p+q)=1.** Si calculamos la entropía, tenemos certidumbre cuando **p= 0,** nunca ocurriría una salida 1 lógico, o cunado **p=1,** es decir nunca ocurría un 0 lógico, a la salida de la fuente de información. En estos dos puntos la FIMN arroja una entropía 0 bits/símbolos. Está claro que la entropía será máxima cuando las probabilidades sean iguales.

$$H(S) = p\ log_2 \frac{1}{p} + q\ log_2 \frac{1}{q} = \frac{bits}{simbolos}$$

16 - Teoría de la información y codificación: El significado de la... SCOZZINA, Fabián Emilio

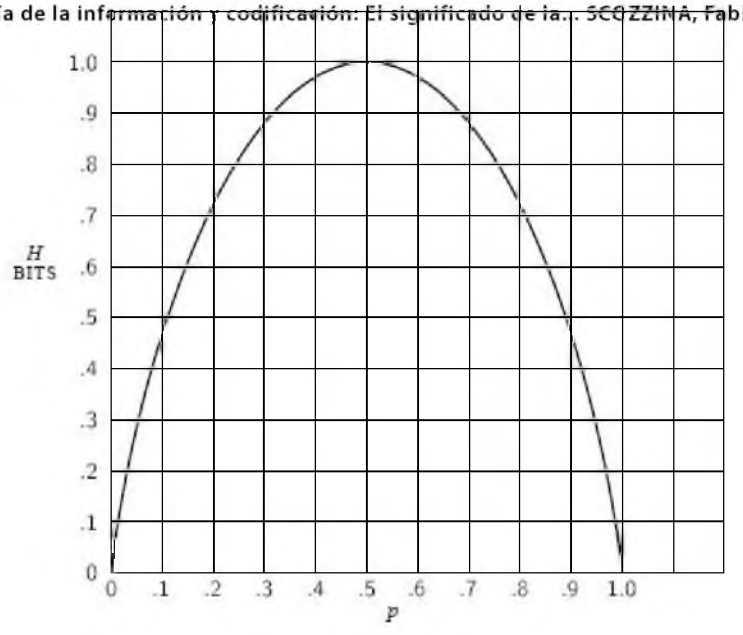


Figura Función entropía con p variable en el intervalo 0 y 1.

La entropía y la estructura del lenguaje.

En su publicación Shannon, describió un experimento, sencillo sobre la estructura del lenguaje. El caso se limita al idioma inglés, tomando 26 letras (símbolos $\mathbf{s}_i$) y un punto, serian en total 27 símbolos, y diseña una serie de fuentes de información, en base a un modelo markoviano.

El experimento consistía en considerar una fuente de información sencilla, al principio sin memoria es decir que todos los símbolos tengan igual probabilidad de ocurrencia P(si)=1/27=0,03703 esta primera *fuente de información,* ***y la llamo aproximación cero.***

Los resultados son una oración sin sentido, inentendible, para el observador de habla inglesa. En las aproximaciones siguientes hizo intervenir las probabilidades reales de los símbolos, según la lengua inglesa, es decir que las letras del alfabeto ya no tienen una misma probabilidad de ocurrencia. Estas aparecen según las reglas de gramática inglés.

Shannon, utilizo varios métodos

para analizar las probabilidades condicionales de que aparezcan en combinadas dos letras, (th,en,ab..); además utilizo fuente de información donde los símbolos eran palabras más comunes en el idioma, ect. Así continúo el experimento, con las probabilidades condicionales, de que aparezcan tres letras (the,end,abs..) Hasta llegar a cinco letras.

Modelo de fuente de información	Entropía Bits/symb	Observaciones
Memoria nula	4,75	Todos los símbolos equiparables, ilegible.
Memoria nula según gramática inglesa	4,03	Lo símbolos tiene una estructura característica de aparición y puntuación, según la lengua inglesa
Markov 1° orden	3,32	Probabilidades de condicionales de Pratt (1942) usadas por Shannon.
Markov 3° orden	3,28	Una estructura similar a la inglesa pero no reconocibles, pero se puede completar en partes la oración.
Markov 4° y 5° orden		Permitia la predicción de la palabras

En nuestra vida diaria usamos de manera asidua, conversores de voz texto, disponibles en las app de los celulares, correctores automáticos de texto, damos órdenes de voz al móvil celular, estos son ejemplos de las aplicaciones industriales de la información.

Como casos de negocios más importantes Amazon Web Services (AWS) es la rama de *cloud computing* de Amazon, para desarrollo de inteligencia artificial. Orienta actualmente los servicios de inteligencia artificial cada vez están más centrados en comprender el lenguaje natural, entender el texto dentro de una conversación del día a día, poder mantener conversaciones usando voz o texto sin problemas de entendimiento e incluso aprender a reconocer caras, objetos y escenas.

Cuando se da la máxima entropía en un alfabeto? El caso particular de una fuente $S=[s_1,s_2,s_3,s_4....s_M]$, con un alfabeto de M símbolos, todos ellos equiparables $P(s_i)=P(s_j)$ nos da una propiedad interesante de la entropía.

$$\sum_{i=1}^{M} P(s_i) = 1$$

$$H(S) = \sum_{i=1}^{M} P(s_i)\, I(s_i) = log_2 \frac{1}{P(M)} \frac{bits}{simbolos}$$

Supongamos una fuente sin memoria, cuando esta fuente emite FIMN[X] símbolos a una tasa **r** bits/símbolos, los cuales presentan distintas probabilidades $P(s_i)$,$P(s_j)$, respecto de otra fuente de información, que emite símbolos con igual probabilidad $P(s_i)=P(s_j)$. Para realizar una transmisión eficiente, se requiere que cada símbolo lleve una cantidad de información diferente, que sea mínima, no haya redundancias y además que sea útil al receptor.

$$\mathrm{R} = \mathrm{rH(X)} < \mathrm{r}\log_2 M$$

El número promedio de dígitos binarios emitidos por cada símbolo de la fuente será:

$$\overline{\mathrm{N}} = \sum_{I=1}^{\mathrm{M}} P_i\, N_i$$

Donde Ni es la longitud del código *i-esimo* y $\mathbf{P}_i$ es la probabilidad de ocurrencia de ese símbolo para transmitirlo. Según el primer teorema de Shannon, establece que la codificación, siempre limitará el valor de dígitos binarios emitidos por cada símbolo.

$$\mathrm{H(X)} \ll \overline{\mathrm{N}} \ll \mathrm{H(X)} + \varepsilon$$

Un código óptimo es aquel que consigue una longitud media **N=H(X)**, en la práctica, la relación de eficiencia del código, es:

$$\eta = \frac{H(X)}{\overline{N}} \leq 1$$

Para que sea posible la decodificación del código, de forma unívoca y no exista perdida de información en el proceso de codificación es necesario que cumpla con la inecuación de ***Krfat***, esta es condición necesaria y suficiente para que el código seleccionado sea decodificable de forma univoca, en el cual las longitudes deben cumplir.

$$K = \sum_{i=1}^{M} 2^{-N_i} \leq 1$$

El código más sencillo de diseñar es aquel en cual todas las palabras tienen la misma longitud.

$$K = M2^{\bar{N}} \leq 1 \; despejando \; N \geq log_2 M$$

$$\eta = \frac{H(X)}{\bar{N}} \leq \frac{H(X)}{logM}$$

Para obtener una eficiencia mayor, hay que usar un código de longitud variable, asignando las probabilidades de aparición más altas a los símbolos de longitud menor, y las más largas a los que tienen menor probabilidad de ocurrencia. Este razonamiento fue aplicado al código Morse para telegrafía donde se usaron para su construcción el idioma inglés.

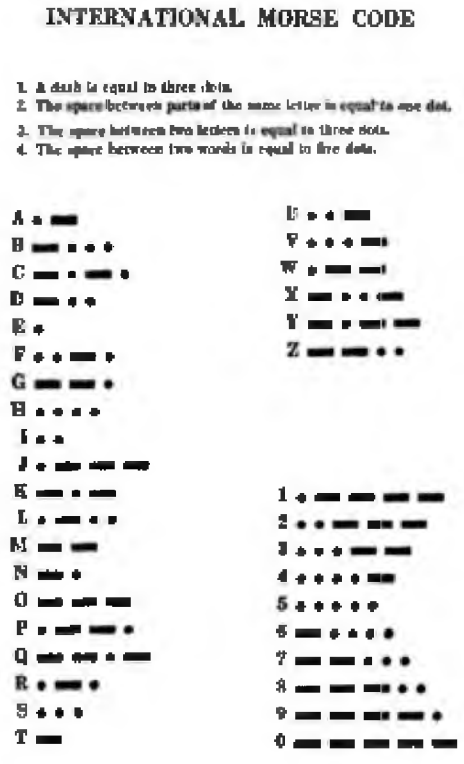


*Figura Código Morse fue desarrollado en 1837 y luego mejorado por Vail, en 1841 al cual fue llamado "**American Morse Code**" y utilizado en la primera transmisión por telégrafo.*

Teniendo en cuenta que la longitud de un código i-estima debe ser entera y será proporcional a la información que lleva el código y estará comprendida entre:

$$I_i \leq N_i \leq I_i + 1$$

$$log\frac{1}{p_i} \leq N_i \leq I_i + 1$$

Al trabajar en un sistema de codificación binaria las probabilidades de los símbolos son potencias de 2. $\mathbf{P_i=2\text{-}N_i}$, donde **i=1,2,3,4... M**. Donde **N**, es número entero.

$$log\frac{1}{p_i} \leq \log\frac{1}{2^{-N_i}} \leq I_i + 1$$

Si analizamos un ejemplo codificación de **Shannon-Fano**, para una fuente de información que tiene cuatro símbolos, con sus correspondientes probabilidades.

Símbolo	Probabilidad	Código SF
X1	0,5	0
X2	0,25	10
X3	0,125	110
X4	0,125	111

Tabla de codificación de Shannon-Fano

Entropía	$H(X) = \sum_{i=1}^{4} P(x_i).\, log\,\frac{1}{P(x_i)} = 1{,}75$
Longitud media	$\overline{N} = \sum_{i=1}^{4} P_i\, N_i = 0{,}5x1+0{,}25x2+0{,}125x3x2 = 1{,}75$
Probabilidad	$P(x_i) = 2^{-N_i} = 0{,}5;\, 0{,}25;\, 0{,}125$
Rendimiento	$\eta = \frac{H(X)}{\overline{N}} = 1$

La entropía y la Información por unidad de tiempo, supongamos que ahora nuestra fuente genera **m** símbolos por segundo. La unidad de **m** es símbolos por segundos (velocidad de generación). El tiempo de producción de símbolos es τ_o= **1/m** y es el máximo que se dispone para generar cada símbolo. Podemos escribir la tasa **R** en función de la entropía.

$$R = \frac{H(S)}{\tau} = mH(S) = \frac{bits}{segundos}$$

Un ejemplo sencillo ocupado en la asignatura para explicar el uso operacional de la entropía en el reconocimiento de imágenes. Se basa en el uso del programa Mathlab, donde se toma una imagen digitalizada, con una con una cámara común. La imagen en BN es tomada al retratar una cartulina negra perforada aleatoriamente sobre un fondo blanco, y luego se calcula la entropía de la imagen con una profundidad de dos bits. Para completar se muestra un histograma de bits 1 y 0, que junto con el valor de la entropía permite reconocer si la imagen cuenta con 15 o menos agujeros.

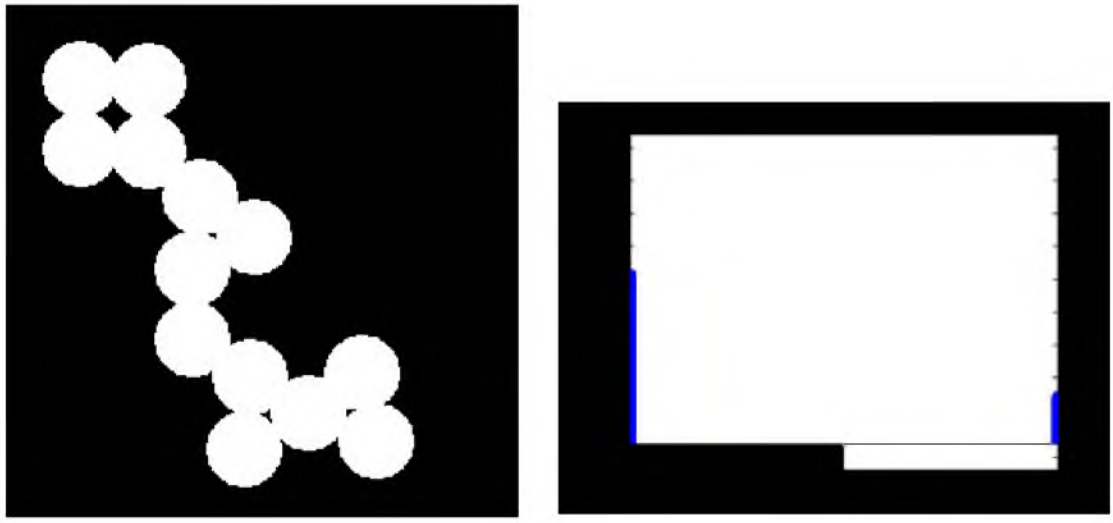


Figura blanco y negro de 256 x 256 pixels, 150 DPI, 4,33x 4,33 cm con una entropía de 0.7522 bits/símbolos. Se muestra también un histograma donde se muestra la cantidad de bits 1 y 0.

Por ultimo transcribiré la histórica conversación entre Claude Shannon y John Von Neumann sobre la entropía. **Shannon comenta**: Pensé en llamarlo "información", pero la palabra se usó demasiado, así que decidí llamarlo "incertidumbre" …. **Von Neumann responde:** "Deberías llamarlo entropía, por dos razones. En primer lugar, tu función de incertidumbre se ha utilizado en mecánica estadística con ese nombre, por lo que ya tiene un nombre. En segundo lugar, y lo que es más importante, nadie sabe qué es realmente la entropía, por lo que en un debate siempre tendrá la ventaja". De este modo Shannon toma la decisión del nombre.

Bibliografía:

Thomas M. Cover, Joy A. Thomas, *"Elements of Information Theory"*, John Wiley & Sons. Second Edition 2006

Zanuy, Fernández, Marcos *Tratamiento digital de voz e imagen y su aplicación a la multimedia.* Alfaomega Marcombo 2001

Skar, Bernad *"Digital Communications Fundamentals and applications"*. Segunda edición en ingles.Prentice Hall PTR 2000

Norman Abramson, *"Teoría de la Información y Codifiación"* Paraninfo Madrid Tercera edición. 1974 Edición original en inglés 1966.

Claude E. Shannon *"A Mathematical Theory of Communication"* published in Bell System Technical Journal in 1948.