Teoría de la información

Antonio Campello [*], Carlos Capote [**]

Resumen

En este documento intentamos hacer comprensibles algunos conceptos básicos de la teoría de la información. Entre otros conceptos, trataremos de esclarecer el significado de *cantidad de información*, *capacidad de un canal*, *entropía* y *ruido*. Tomaremos estos conceptos en el mismo sentido en que fueron usados en los documentos que sentaron las bases de la teoría de la información: *Transmission of Information* de R. V. L. Hartley (1928) y *A Mathematical Theory of Communication* de C. E. Shannon (1948). Trataremos de hacer de este documento una introducción a la teoría de la información accesible a cualquier persona sin conocimientos previos sobre el tema[1].

Palabras clave: cantidad de información, capacidad, entropía, ruido

1. Introducción

1.1. Esquema de comunicación

Todo el mundo ha oído hablar en algún momento de algún modelo básico de comunicación. Estos modelos constan siempre, al menos, de un emisor, un receptor, un mensaje y un canal. Y es que, aún siendo un modelo un poco precario, es más que suficiente para explicar con un grado aceptable de detalle casi cualquier proceso de comunicación. En general, en los procesos que describimos con estos modelos tan simplificados, la comunicación es bidireccional: emisor y receptor pueden intercambiar sus papeles. De esta manera, aún con un esquema tan simple, podemos describir aceptablemente el funcionamiento de la televisión, la radio, el teléfono, el correo postal, la conexión entre el ordenador y la impresora, entre el teclado y la placa base del ordenador, la comunicación entre un pastor y sus ovejas o incluso entre una neurona y otra.

A Shannon, un ingeniero electricista y matemático conocido por sentar las bases de la teoría de la información, debemos el esquema de la Figura 1. El esquema de Shannon es sustancialmente el mismo esquema del que hablábamos. Este esquema puede entenderse fácilmente con un ejemplo cotidiano como el envío de un mensaje de texto (SMS) a través del teléfono móvil. La fuente de información y el destino, en este caso, son la persona que envía y la que recibe el mensaje respectivamente. Transmisor y receptor son los teléfonos móviles, que son capaces de transformar el texto en una señal electromagnética en el origen y de realizar el proceso inverso en destino para recuperar el mensaje.[2] El origen de ruido representa todo aquello que pueda perturbar la

[*] Correo electrónico: accampellojr@gmail.com

[**] Correo electrónico: mail@carloscapote.com

[1] Supondremos, eso sí, que se dominan algunos conceptos matemáticos propios de la educación preuniversitaria, como las propiedades de los logaritmos o la diferencia entre una función exponencial y una lineal, determinante de una matriz, probabilidad condicional, representaciones de Markoff, etc ;-)

[2] No es necesario recordar que esta aplicación del esquema es también una simplificación en la que obviamos, entre otros, el papel de las redes telefónicas.

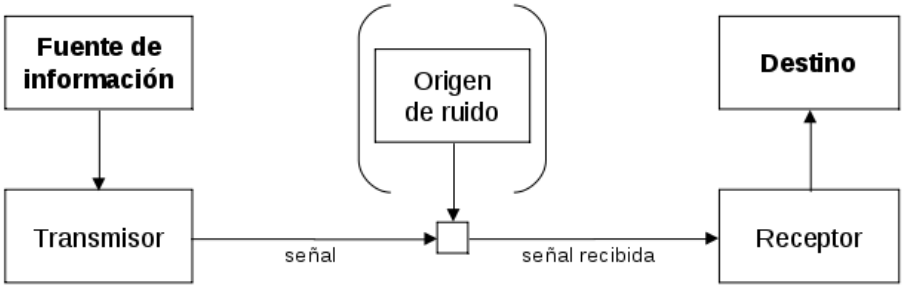

Figura 1: Diagrama esquemático de un sistema general de comunicación

señal enviada, a lo largo de su camino hasta llegar al destino, de manera que la señal recibida no sea una réplica idéntica de la señal original.

Ahora es cuando, haciendo un breve repaso pensamos: de acuerdo, entiendo que el esquema de un sistema de comunicación representa de forma abstracta cualquier proceso posible de envío de información, ¡pero aún no sé qué es información!

1.2. ¿Qué es información?

Intuitivamente, diremos que información es cualquier cosa que puede hacer llegar un emisor a un receptor. En concreto, es algo que viaja con lo que envía el emisor y que el receptor es capaz de reconstruir. Podríamos liarnos a decir que el mensaje debe tener sentido para emisor y receptor, y tratar de encontrar una definición formal de información en ese *sentido* pero estaríamos desviándonos completamente del rumbo de la teoría de la información.

Hamming, 1980
"La teoría de la información no trata el significado de la información, sólo trata la cantidad de información"

Como decía Hamming, en teoría de información lo que nos interesa es la cantidad de información, no su significado. Sin embargo, que no nos interese el significado de la información enviada no debería implicar que no nos interese el significado del concepto información. No obstante, como en tantas otras ocasiones, para definir el concepto información puede ser suficiente con lograr definir una medida de la cantidad de información. Y eso es lo que haremos.

1.2.1. El código Simpson

En un episodio de los Simpsons, Abraham Simpson, el abuelo, espera en una barca que acaban de tomar prestada a la familia Flanders mientras Bart bucea en busca de una caja muy valiosa. Justo antes de que Bart se sumerja establecen entre ellos un código para comunicarse a través de una cuerda. Si Bart se queda sin aire debe tirar 63 veces de la cuerda. Si encuentra el tesoro debe tirar 64 veces. Si Shannon hibiese estado allí, les habría frito a collejas.

Shannon, 1948
"Los aspectos semánticos de la comunicación son irrelevantes al ingeniero. Lo significativo es que el mensaje en cuestión es *seleccionado de un conjunto* de posibles mensajes"

Abraham y Bart querían establecer un medio de comunicación donde sólo dos mensajes tenían interés. Su diccionario, dicho de otra manera, sólo tenía dos palabras. Para más inri, la comunicación sólo tenía sentido en un sentido (de Bart a Abraham) y no podían crearse secuencias a partir de los mensajes. ¡La comunicación debía limitarse a enviar uno de los dos mensajes posibles!

El código Simpson era realmente malo porque fallaba de lleno en la codificación. Al elegir cómo codificarían cada uno de los mensajes posibles sólo tenían que haber tenido en cuenta dos ideas.

En primer lugar, los códigos asignados a los dos mensajes deben ser fácilmente distinguibles. En los laboratorios se enseña a los científicos experimentales a trabajar con incertidumbres asociadas a cada medida. Si un aprendiz de científico estuviese en el lugar de Abraham Simpson pensaría: "soy capaz de distinguir un tirón de cuerda de dos y de ninguno pero no de medio tirón de cuerda o de uno y medio"de modo que asignaría una incertidumbre de a cada medida realizada. Así, en el mejor de los casos la medida asociada al mensaje *me he quedado sin aire* sería $m_1 = 63 \pm 1$ tirones de cuerda, y la asociada al mensaje *he encontrado el tesoro* sería $m_2 = 64 \pm 1$ tirones. ¡Pero ambas medidas son casi indistinguibles!

En segundo lugar, una de las funciones de la codificación en fuente es encontrar una manera compacta de enviar el mensaje. ¡Pero los primeros 62 tirones de cuerda no ayudan en absoluto a distinguir entre un mensaje y otro! Hacer más largo un mensaje puede tener sentido cuando se haga para añadir redundancia a la señal, protegiéndola del ruido y facilitando la reconstrucción del mensaje en destino... pero evidentemente no es el caso.

1.3. Medir la información

Hartley, que era otro tipo listo, pensó que una medida de información debía tener algún tipo de relación con el número de mensajes posibles. Una persona que habla, por ejemplo, va escogiendo palabras para construir un mensaje. En "Hartley era investigador"la primera palabra elimina a todas aquellas personas que no son Hartley y cualquier otro tipo de cosa que no se llame Hartley en general, la segunda dirige la atención hacia un atributo de Hartley y la tercera elimina otras posibles profesiones.

De esta manera, según aumente el número de símbolos disponibles en un determinado acto de comunicación, aumentará también el número de mensajes posibles. De hecho, es posible modelar un sistema de comunicación que nos permita estudiar en qué medida aumenta el número de mensajes posibles cuando se añade una palabra a un mensaje.

En el código Simpson sólo habían dos mensajes posibles y no se contemplaba la posibilidad de enviar sucesiones de mensajes. Sin embargo, por lo general, en un sistema de comunicación dispondremos de un conjunto de símbolos y podremos enviar sucesiones de ellos. Supongamos que disponemos de 2 símbolos: 0 y 1 (o punto y raya) y que podemos escoger 3 de esos símbolos de la manera que queramos. Es fácil ver que podríamos generar $2^3 = 8$ mensajes posibles. Si en lugar de 3 pudiésemos escoger un número indeterminado n de símbolos el número de mensajes que podríamos generar es 2^n. Por otro lado, si en lugar de 2 símbolos dispusiésemos de s símbolos el número de mensajes que podríamos generar es s^n.

Para que una medida de información tenga valor práctico desde el punto de vista de la ingeniería debe variar linealmente con el número de símbolos seleccionados en un mensaje. Así, aprovechamos la magnífica propiedad:

$$log_a b^n = n log_a b \tag{1}$$

De esta manera, hemos conseguido lo que queríamos. ¡Hemos llegado a una medida práctica de información! La medida de información que debemos a Hartley es:

$$I = log s^n \tag{2}$$

Es decir, hemos tomado como nuestra medida de información el logaritmo del número de posibles secuencias de símbolos.

1.3.1. El telégrafo

Aunque hemos avanzado mucho, nuestra medida de información todavía necesita algunos matices. De hecho, ni siquiera un ejemplo tan simple y clásico como el de la información transmitida por un telégrafo puede ser estudiada convenientemente mediante la definición anterior. Veamos porqué.

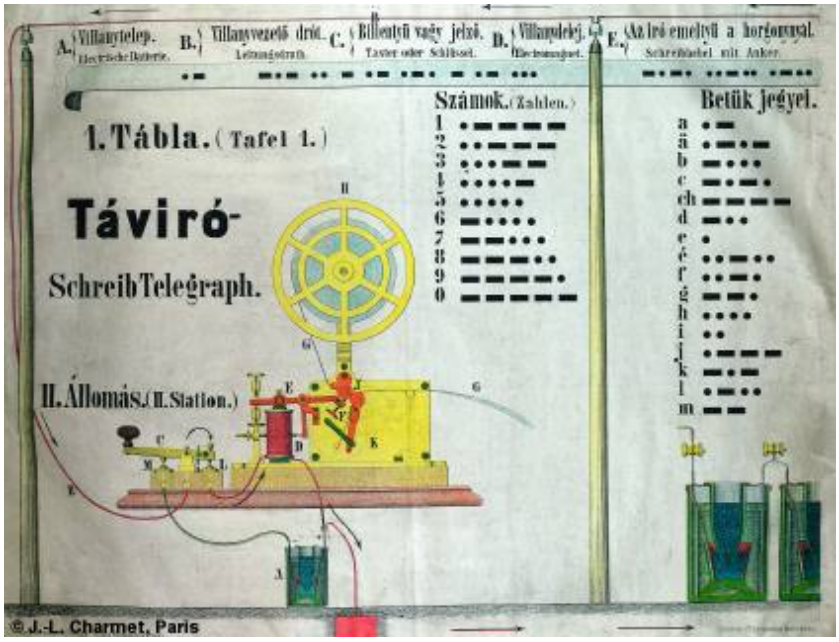

Figura 2: Código Morse

Como es sabido el telégrafo funciona enviando puntos y rayas. Supongamos, por un momento, que a cada combinación de 3 puntos y rayas se le hiciese corresponder una letra del alfabeto[3]. Hemos dicho que nos gustaría que, por comodidad, la medida de la información contenida en 3 caracteres fuese el triple que la información contenida en 1 caracter.

...	- - -	...
S	O	S

Cuadro 1: Ejemplo simplificado de codificación

Si llamamos n_1 al número de símbolos primarios (puntos y rayas) por cada símbolo secundario (caracter) y n_2 al número de caracteres secundarios que queremos seleccionar, vemos que se sigue cumpliendo la relación de linealidad que esperábamos:

$$I_t = logs^{n_1 n_2} = n_2 logs^{n_1} \qquad (3)$$

Pero, para variar, ¡el telégrafo no es tan simple! Si quisiésemos[4] codificar el alfabeto en puntos y guiones necesitaríamos al menos 5 símbolos ($2^5 = 32$). Sin embargo, la codificación morse no utiliza más de 4 puntos y rayas por caracter. ¿Cómo lo hace?

En la Figura 2 se aprecia que el código Morse no asigna el mismo número de puntos y rayas a cada letra. Alguien tuvo la idea, genial idea, de hecho, de asignar a las letras más usadas en inglés los códigos más cortos posibles en Morse y viceversa. Así, la E y la T, que son las letras más usadas en inglés, se codifican con un punto y con un raya respectivamente. ¡Este es un buen ejemplo de codificación en la fuente!

Y ahora es cuando vienen los nuevos *peros*: ¡Ahora dos conjuntos de n caracteres no tienen porqué contener la misma cantidad de información! Es más, ¡cómo podemos distinguir la letra S, que se codifica como 3 puntos, de la sucesión de 3 letras E!

La respuesta a estas preguntas vendrá por partes. En primer lugar, la cantidad de información será lineal en el número de símbolos primarios (es decir, de puntos y rayas, ceros

[3] ¡Ya! ¡Lo sé! ¡Sólo podríamos codificar 8 de las letras del alfabeto!

[4] De hecho, el teletipo es un caso particular de telégrafo que funciona justo de esta manera: siempre se cumple la correspondencia 5 bits / símbolo.

y unos, etc) pero sólo lo será en el número de caracteres si la relación entre caracteres y símbolos primarios es constante. En segundo lugar, necesitaremos añadir restricciones a nuestra codificación para evitar problemas como el de confundir la letra S con tres E's seguidas.

El problema de las restricciones se podría resolver fácilmente en este caso[5], pero de la imposición de restricciones nacen nuevas dificultades. Veamos un ejemplo de restricción real pero antes, hagamos un repaso del funcionamiento del telégrafo.

Imaginemos un telégrafo como una manguera por la que podemos hacer pasar agua. Por convenio establecemos una unidad de tiempo (nos da igual de qué unidad de tiempo se trate aunque, cuanto menor sea, más rápida será la transmisión de información). Lo que hasta ahora hemos llamado *punto*, consistirá en mantener el grifo cerrado durante una unidad de tiempo y abierto la siguiente. La transmisión de un punto llevará, entonces, dos unidades de tiempo. Lo que llamábamos *raya* consistirá en tres unidades de tiempo con el grifo cerrado y una unidad de tiempo con el grifo abierto. Introduciremos también un espacio entre letras como tres unidades de tiempo con el grifo abierto y un espacio entre palabras como seis unidades de tiempo con el grifo abierto.

Punto	0	1				
Raya	0	0	0	1		
L-Espacio	1	1	1			
P-Espacio	1	1	1	1	1	1

Cuadro 2: Símbolos disponibles en el telégrafo
1: línea abierta; 2: línea cerrada

Una restricción cuyo estudio es sencillo es la que impide el envío de dos espacios seguidos (ya sean espacios entre letras o entre palabras). Esto implica que en el telégrafo pueden darse dos estados: 1) acabamos de enviar un espacio y, por tanto, sólo podemos enviar un punto o una raya ó 2) acabamos de enviar un punto o una raya y por tanto podemos enviar cualquiera de los símbolos disponibles (ver Figura 3).

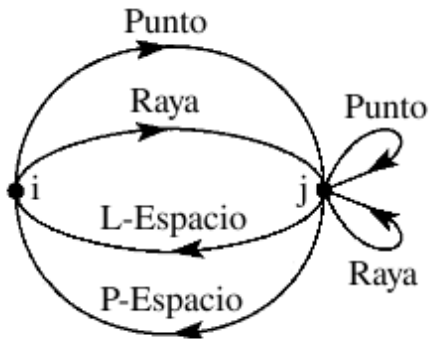

Figura 3: Restricciones en el código Morse

2. Canales sin ruido

2.1. Capacidad

Uno de los objetivos principales de la teoría de la información es determinar la capacidad de diferentes canales. Siguiendo nuestra línea habitual, pongamos primero un ejemplo intuitivo de lo que es la capacidad de un canal. Imaginemos un enorme bidón de agua con una pequeña manguera en su base por la que el agua se puede escapar. Si contamos los litros de agua que salen del bidón cada minuto, seguro que obtendremos una cantidad mucho menor que la que obtendríamos si la manguera fuese mucho más grande.

El teletipo es un caso particular de telégrafo en que todos los símbolos tienen la misma duración. El teletipo puede trabajar con 32 ($= 2^5$) símbolos, donde cada símbolo representa 5 bits de información. Si un teletipo transmite n símbolos por segundo, es fácil ver que estará trabajando a $5n$ bits por segundo. La definición de capacidad para un canal discreto es:

$$C = \lim_{T\to\infty} \frac{\log N(T)}{T} \qquad (4)$$

donde $N(T)$ es la cantidad de símbolos de

[5]Bastaría con prohibir el envío de dos o más E's seguidas.

duración T que permite el canal y C, usualmente se mide en $bits/seg$.

2.1.1. Representaciones

Es posible probar que el límite que define la capacidad de un canal discreto es finito en muchos casos. Sin embargo, encontrar su valor puede ser muy difícil. Una manera de simplificar este cálculo es a través de las representaciones gráficas de los posibles estados del sistema. Volviendo a la Figura 3 recordamos que en el caso del telégrafo tenemos 2 posibles estados i y j.

Si llamamos $b_{ij}^{(s)}$ a la duración del símbolo s que lleva el estado de i a j, entonces[6] la capacidad del canal vendrá dada por $logw$, con w la raíz real más grande del polinomio dado por el determinante de:

$$A_{ij} = \sum_s W^{-b_{ij}^{(s)}} - \delta_{ij} \tag{5}$$

Siguiendo con el ejemplo del telégrafo, vemos que ningún símbolo puede llevar el estado i al estado i. Traducido al lenguaje del teorema:

$$\sum_s W^{-b_{11}^{(s)}} = 0 \tag{6}$$

Sin embargo, tanto el punto (de duración 2) como la raya (de duración 4) pueden llevar el estado j al estado j. En el lenguaje del teorema:

$$\sum_s W^{-b_{22}^{(s)}} = W^{-2} + W^{-4} \tag{7}$$

Operando de este modo para el resto de casos, llegamos a la expresión de la ecuación que determina la capacidad del telégrafo:

$$\begin{vmatrix} -1 & W^{-2} + W^{-4} \\ W^{-3} + W^{-6} & W^{-2} + W^{-4} - 1 \end{vmatrix} = 0 \tag{8}$$

de donde estimamos la capacidad del telégrafo en: $C = 0,5389$ bits / ud. de tiempo.

2.2. Entropía

Como veremos más adelante nos será muy útil disponer de una magnitud que nos dé una idea de la *sorpresa* asociada a los posibles resultados de un determinado proceso. Pero antes de dar una definición formal de entropía, veamos una definición intuitiva y algunas de las propiedades que queremos que cumpla esta función.

2.2.1. Definición intuitiva

Imaginemos que tenemos una moneda perfectamente calibrada, de tal manera que la probabilidad de que al lanzarla caiga de cara es exactamente 1/2 y la de que caiga cruz es... también de 1/2 (vamos, que no cae de canto ni de casualidad). Diremos que la sorpresa asociada a este proceso es máxima, dado que no tenemos ni la menor idea de qué va a pasar. Imaginemos ahora que trucamos la moneda de tal manera que la probabilidad de que salga cara es de 99/100 y la de que salga cruz es de un escaso 1/100. En esta ocasión la sorpresa asociada al proceso es mucho menor.[7] Lo que buscamos es, por tanto, algo que será función de las diferentes distribuciones de probabilidad.

2.2.2. Propiedades

Si, en el caso de la moneda, propusiésemos una distribución de probabilidades de 98/100 frente a 2/100, esperaríamos que la nueva función tuviese un comportamiento muy parecido al que tenía en el caso 99/100 frente a 1/100. En términos algo más formales:

- H debe ser **continua** en p_i.

Hasta ahora hemos comparado diferentes distribuciones de probabilidad para casos en que sólo dos eventos de probabilidades p y $q = (1 - p)$ eran posibles. Veamos ahora

[6] Consultar el paper de Shannon para encontrar la demostración formal.

[7] ¡Ojo! Lo que nos interesa es una medida de la entropía del proceso. Ya sabemos que si saliese cruz la sorpresa sería muy grande, pero esta sorpresa se referiría a un evento concreto, y nosotros estamos estudiando la sorpresa de una manera más "global", referida al proceso en general.

cómo querríamos que se comportase la nueva función cuando la distribución de probabilidades ser similar y la diferencia venga impuesta por el número de eventos posibles. Parece claro que los resultados posibles de arrojar una moneda al aire llevan asociada una sorpresa menor que los resultados posibles de lanzar un dado. Dicho en otras palabras:

- Si las probabilidades de dos procesos son siempre idénticas y de la forma $p_i = 1/n$, entonces H debe ser una función **monótona creciente** de n.

Las dos propiedades anteriores determinan cómo debe comportarse la función cuando varía la distribución de probabilidades para un mismo número de eventos, y cuando varía el número de eventos para una misma distribución. Exijamos ahora que:

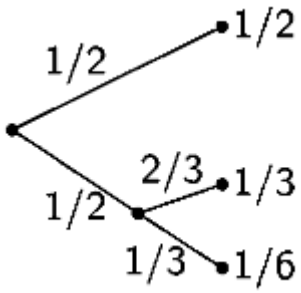

Figura 4: Descomposición de decisiones

- Si una elección se rompe en varias elecciones, la H original sea la suma ponderada de valores individuales de la nueva H.

2.2.3. Definición formal

Puede demostrarse[8] que la única H que satisface las tres exigencias anteriores es de la forma:

$$H = -K \sum_{i=1}^{n} p_i \log p_i \qquad (9)$$

Que, particularizada para un caso como el de la moneda, es decir, de dos probabilidades p y $q = 1 - p$, queda en la forma:

$$H = -(p \log p + q \log q) \qquad (10)$$

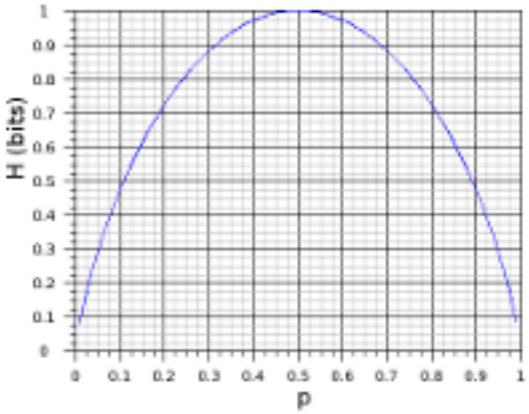

Figura 5: Entropía en el caso de 2 posibilidades de probabilidades p y 1-p

2.2.4. Medida de información

La entropía juega un papel fundamental en la medida de la información. Recordemos que cumple algunas propiedades que la hacen especialmente interesante para medir información:

1. $H = 0$ sólo cuando las distribuciones de probabilidades son del tipo *todo ceros* y un sólo evento con una probabilidad 1. Esto coincide con la idea que tenemos de información porque, dicho de otra manera, si un evento está completamente determinado y sabemos perfectamente qué va a pasar, entonces que suceda no nos aporta nada de información.

2. Para un número de eventos dado, la entropía es máxima (e igual a $\log n$) cuando las probabilidades son todas iguales ($p_i = 1/n$).

3. De hecho, se puede demostrar que dadas las probabilidades de dos eventos p_1 y p_2

[8] Ver apéndice 2 del paper de Shannon.

de tal manera que $|p_1 - p_2| = \delta > 0$, si hacemos cualquier cambio que haga más pequeña la diferencia δ entre ambas probabilidades provocará un aumento en la entropía[9].

Las medidas de información que habíamos visto hasta ahora eran buenas siempre y cuando se diesen ciertas condiciones. Eran medidas, en cierto modo, absolutas. La entropía es una medida de la cantidad media de información asociada a una variable aleatoria.

2.2.5. Entropía condicional

Las propiedades que exigimos en su momento para definir la entropía nos permiten ahora definir una entropía condicional que, por analogía con la probabilidad condicional, definiremos como la entropía de un evento considerando que se conoce el resultado de otro.

$$H_x(y) = -\sum_{i,j} p(i,j) \log p_i(j) \qquad (11)$$

Esta es una medida promedio de cuánta incertidumbre tenemos sobre y cuando conocemos x. Sustituyendo algunos términos se llega fácilmente a:

$$H(x,y) = H(x) + H_x(y) \qquad (12)$$

2.2.6. Entropía de una fuente

La entropía de una fuente viene dada por la expresión:

$$H = \sum_i P_i H_i \qquad (13)$$

Pensemos un momento en lo que significa la entropía de una fuente. Hasta ahora hemos dicho que la entropía era una medida de la información que se basaba en la incertidumbre asociada a una determinada variable aleatoria. Imaginemos, entonces, que la fuente de un sistema de comunicación es un telegrafista y que quiere enviar un mensaje en español. ¿Será cero o máxima la incertidumbre de la fuente? Seguramente ni lo uno ni lo otro. Si la incertidumbre de la fuente fuese cero sabríamos perfectamente qué nos va a enviar y no es el caso, pero tampoco será máxima porque la sucesión de caracteres que se enviarán no será completamente aleatoria. Una secuencia dada de caracteres puede llegar a determinar completamente el siguiente caracter. Por ejemplo, dada la secuencia de caracteres: *Shannon es considerado padre de la teoría de la informació* es mucho más probable que el siguiente caracter sea una n que cualquier otro caracter. Hasta el punto de que si recibimos otro caracter probablemente consideremos que se trata de un error y pongamos la n en su lugar.

¿Esto significa que los lenguajes hablados (español, inglés, etc) tienen en sus propias tripas incorporada cierta redundancia? Definitivamente sí. Hasta el punto que Shannon, en su paper llega a cuantificar la redundancia del inglés en un 50 %, lo que significa que si eliminamos aleatoriamente la mitad de los caracteres de un texto en inglés alguien con buenos conocimientos del idioma debería de ser capaz de reconstruirlo.

2.3. Teorema fundamental

Lo que hemos visto hasta ahora nos da la idea de que a la hora de enviar información a través de un determinado canal debemos tener en cuenta cómo codificarla. Dicho de otra manera, hemos visto que podemos comprimir un mensaje desde la fuente aún sin perder nada de información. Como vimos en el caso del telégrafo, donde se asignaban códigos más cortos a las letras más usadas en inglés. También hemos visto que podemos agregar redundancia a un mensaje haciendo que pueda reconstruirse si se pierde una determinada parte por el camino. El teorema fundamental de Shannon para canales sin ruido propone que:

Dada una fuente de entropía H (bits por símbolo) y un canal de capacidad C (bits por segundo). Entonces, es posible encodificar la

[9] La entropía aumenta con cualquier operación de *averaging*.

salida de la fuente de tal manera que se transmita a una tasa de transmisión $\frac{C}{H} - \epsilon$ símbolos por segundo sobre el canal, donde ϵ es arbitrariamente pequeño. No es posible transmitir a una tasa de transmisión superior a $\frac{C}{H}$.

Una manera de entender este teorema es imaginar que un profesor coloca a cuatro o cinco de sus alumnos formando un teléfono humano. El primero de ellos estará en contacto con Alice, la fuente, y el último estará en contacto con Bob, el destinatario del mensaje. El profesor, que es muy metódico, se toma la molestia de determinar la capacidad del canal, estimándola en, por ejemplo, 2 bytes / segundo. Intuitivamente, tendemos a pensar que ya está plenamente determinada la cantidad de información que se puede enviar a través de este canal. ¿Qué pasa si Alice habla muy rápido? Lo normal es pensar que de ninguna manera íbamos a lograr que la información atravesase íntegramente el canal. No obstante, según este teorema todo depende de lo loca que esté Alice. Si su discurso es completamente incoherente y usa palabras inventadas, o concatena palabras sin sentido alguno, probablemente el canal no de abasto. En este caso, la entropía de Alice sería máxima y la cantidad de información (C/H) que podríamos enviar a través del canal se haría mínima. Si, en caso contrario, el discurso de Alice relativamente predecible, la entropía de Alice se hará mínima y la cantidad de información que podremos enviar a través del canal se maximizará.

3. Canales con ruido

3.1. Tasa de transmisión

En la era pre-Shannon un canal con ruido era algo bastante mal visto. Si un canal tenía ruido, prácticamente lo único que se podía hacer era: 1) revisar el circuito ó 2) enviar el mensaje muchas veces para asegurar que llegaría correctamente. En la era post-Shannon, sabemos que existen maneras de codificar un mensaje que permiten alcanzar tasas de transmisión máximas. Pero antes, vamos a ver algunas complicaciones que nos llevan a tener que redefinir nuestra manera de medir la tasa de transmisión en un canal con ruido.

Supongamos que podemos enviar dos símbolos, 0 y 1, con probabilidades iguales: $p_1 = p_2 = 1/2$. Y supongamos que enviamos a una tasa de 1000 símbolos por segundo. Si hay presente ruido en el canal y provoca que 1 de cada 100 símbolos lleguen mal, ¿cuál será nuestra tasa de transmisión? La intuición nos dice que, como 10 de 1000 símbolos llegan mal y en un segundo enviamos 1000, la tasa de transmisión ha de ser 990 símbolos por segundo. ¡Pero es incorrecta! Esta manera de operar no tiene en cuenta que el destinatario desconocerá dónde se han producido los errores.

Ya vimos que la entropía era una medida del ratio de producción de información de una fuente y estudiamos el concepto de entropía condicional, de modo que no nos extrañará demasiado la definición de tasa de transmisión que propone Shannon:

$$R = H(x) - H_y(x) \tag{14}$$

Si volvemos al ejemplo, tenemos que $H(x) = 1$ bit / símbolo de manera que nos queda por determinar $H_y(x)$. Volviendo a la definición de entropía condicional tenemos que:

$$\begin{aligned} H_y(x) &= -(0,99 \log 0,99 + 0,01 \log 0,01) \\ &= 0,081 \text{ bits / símbolo} \end{aligned}$$

ó 81 bits por segundo. De esta manera, podemos decir que, en el ejemplo, la tasa de transmisión es de $1000 - 81 = 919$ bits / segundo. En el caso extremo de que en destino los 0's enviados sean recibidos con la misma probabilidad como 0's ó 1's (e ídem para los 1's), las probabilidades en destino serían 1/2 y la entropía condicional sería:

$$\begin{aligned} H_y(x) &= -(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}) \\ &= 1 \text{ bit / símbolo} \end{aligned}$$

ó 1000 bits por segundo. De manera que la tasa de transmisión en este caso (caso de máximo ruido) es de $1000 - 1000 = 0$ bits / segundo.

3.2. Capacidad

Una vez definida la tasa de transmisión para un canal con ruido resulta relativamente sencillo proponer una definición de capacidad. La capacidad debe ser el máximo ratio posible de transmisión:

$$C = Max(H(x) - H_y(x)) \qquad (15)$$

3.3. Teorema fundamental

Sea un canal discreto de capacidad C y una fuente discreta de entropía (por segundo) H. Si $H \leq C$ existe un sistema de codificación tal que la salida de lafuente puede ser transmitida con una frecuencia de errores arbitrariamente pequeña.

Dicho en castellano, esto quiere decir que: ¡se pueden corregir errores prácticamente a la perfección si el ratio de producción de información de la fuente es inferior a la capacidad de un canal! Lo malo (o lo bueno, según se mire) es que el teorema de Shannon es no constructivo y sólo nos dice que existen esos códigos, ¡pero no nos dice cómo encontrarlos!

Referencias

[1] R. V. L. Hartley, *Transmission of Information*, 1928

[2] C. E. Shannon, *A Mathematical Theory of Communication*, 1948