Teoría de la información y codificación sensorial

Daniel Herrera

August 28, 2018

1 Introducción

La teoría de la información es el estudio matemático de los procesos de codificación y transmisión de la información. Esta teoría, publicada en 1948 por el matemático Claude Shannon en los laboratorios Bell[1], fue uno de los pilares del desarrollo de las telecomunicaciones en el siglo XX y de la revolución electrónica. Además de ser utilizada para estudiar procesos físicos e ingenieriles, esta teoría ha sido utilizada para estudiar sistemas biológicos, especialmente el sistema nervioso.

El rol que ha tenido esta teoría en el estudio de estos sistemas biológicos es doble. Por un lado, al ser una teoría general de la de la transmisión y codificación de la información, permite elaborar hipótesis unificadoras y exactas sobre los procesos de manipulación de información que lleva a cabo el sistema nervioso, el órgano procesador de información por excelencia. Por otro lado, es una herramienta que nos permite realizar descripciones cuantitativamente precisas de fenómenos de interés biológico.

Con respecto al su rol como generador de hipótesis normativas: En las áreas de la ciencia dedicadas al estudio de sistemas complejos es importante la existencia de teorías subyacentes que generen sentido de la gran cantidad de datos empíricos que de otra forma puede parecer dispares, y que permitan distinguir entre los aspecto más y menos importantes de los mismos. El neurocientífico británico Horace Barlow ilustra este pensamiento en 1961 con la siguiente analogía:

> Un ala sería una estructura sumamente mística si uno no supiera que las aves vuelan. Uno podría observar que se extiende una distancia considerable, que tiene una suave cobertura de plumas con marcas llamativas, que es operada por poderosos músculos, y que la liviandad y la fuerza son aspectos prominentes de su arquitectura. Estos son hechos importantes, pero no nos dicen por sí mismos que las

[1] Los laboratorios Bell fueron fundados por el inventor del teléfono, Alexander Graham Bell, y eran manejados en aquel entonces por la empresa de telfonía AT&T. 8 premios nobel se otorgaron por trabajo realizado en estos laboratorios

> aves vuelan. Sin embargo, sin entender esto, y sin entender algo de los principios del vuelo, un examen más detallado del ala en sí probablemente no sería recompenzante. Creo que estamos en un punto analogo en nuestro entendimiento de la parte sensorial del sistema nervios. Contamos con la primera tanda de hechos provenientes del estudio anatómico, neurofisiológico y psicofísico de la sensación y la percepción, y ahora necesitamos ideas sobre qué operaciones llevan a cabo las varias estructuras que hemos examinado. Sobre el ala del ave puede decirse que acelera el aire hacia abajo transversándolo, y así genera una fuerza hacia arriba que soporta el peso del ave; cual podría ser el resumen similar de la operación más importante realizada por un área sensorial?

En estas líneas, dado que los sistemas sensoriales cumplen el rol general de procesar y transmitir la información que obtienen del entorno para generar acciones, podemos utilizar la teoría de la información para hacer una mejor especificación de los procesos que pueden estar llevando a cabo.

Con respecto al rol de la teoría de la información como herramienta en la neurociencia: Es difícil poner a prueba de forma adecuada muchas hipótesis o preguntas interesantes con respecto al procesamiento de información si no contamos con las herramientas cuantitativas adecuadas. Una fuente de preguntas interesantes en este aspecto es la investigación de los códigos neurales.

Las neuronas generan trenes de potenciales de acción (PA) en respuesta a las señales que reciben de otras neuronas o de los receptores sensoriales. Estos trenes de PA son el medio que las neuronas utilizan para comunicarse entre ellas, pero es importante preguntarnos, ¿qué código utilizan en esos trenes?, ¿qué lenguaje están hablando?, ¿qué información sobre el mundo están transmitiendo en esos potenciales de acción?, ¿cuanta información pueden extraer de esos trenes de PA las neuronas que los reciben?, ¿qué tan confiable es la información que transmite una neurona?. Estas y otras preguntas requieren de alguna forma u otra hacer preguntas precisas sobre la información en estos códigos, lo que requiere formular las preguntas en términos matemáticos, y cuantificar la información contenida en los trenes de PA observados empíricamente. Esto es lo que nos permite hacer la teoría de la información.

Poco después de que Shannon formule la teoría de la información, se reconoció su potencial importancia para varias áreas científicas, entre ellas la neurociencia y la psicología. Desde entonces y hasta hoy (quizás pueda decirse que este enfoque actualmente se encuentra en auge) la teoría de la información ha sido utilizada en estas áreas. A continuación haremos un breve repaso de los conceptos fundamentales de la teoría de la información, y luego ilustraremos brevemente dos aplicaciones que ha tenido en el estudio de los sistemas sensoriales.

2 Incertidumbre, sorpresa e información

El objeto de estudio de la teoría de la información es lo que coloquialmente nos referimos como información, y la misma da definiciones precisas y expande conceptos que ya nos son más o menos intuitivos. Uno de los conceptos principales en la teoría es el de la incertidumbre, o la sorpresa, que puede ser formalizado mediante el uso de la teoría de probabilidad.

Comenzamos explicando de forma medianamente formal qué es la información y luego pasamos a ejemplos intuitivos.

La información es aquello que obtenemos al observar un evento o recibir un mensaje, de forma que nuestra incertidumbre sobre el mundo se ve reducida. De hecho podemos pensar en la información directamente como reducción en la incertidumbre.

Pensemos en que vamos a observar una variable aleatoria A, que puede tomar diferentes valores $\{a_1, a_2, \ldots, a_n\}$, cada uno con probabilidad $\{\mathsf{P}(a_1), \mathsf{P}(a_2), \ldots, \mathsf{P}(a_n)\}$[2]. Antes de observar el valor que tomó la variable A, tenemos incertidumbre con respecto a qué valor tiene o tendrá la misma. Luego de observar qué valor tomó, nuestra incertidumbre se redujo, ganamos información. La información obtenida al observar que una variable aleatoria A toma un determinado valor a_i se define de la siguiente forma:

$$\mathsf{I}\left(A = a_i\right) = -\log_2(\mathsf{P}(a_i))$$

donde si utilizamos el logaritmo en base 2 como aparece en la fórmula, la información está en unidades de bits.

Comencemos con los ejemplos clásicos de probabilidad:

(1) Tiraremos una moneda y observaremos el resultado. En este caso la variable aleatoria A es la tirada de la moneda, y los posibles valores a_i son $\{Cara, Cruz\}$, con probabilidades $\{\mathsf{P}(Cara) = 0.5, \mathsf{P}(Cruz) = 0.5\}$. En este caso la información obtenida al observar que la moneda cayó con *Cara* hacia arriba es[3]:

$$\mathsf{I}(A = Cara) = -\log_2(0.5) = 1 \text{ bit}$$

(2) Hacemos un experimento similar tirando un dado. La variable aleatoria es el resultado de la tirada, y los posibles valores que puede tomar son los números del 1 al 6, cada uno con probabilidad $\frac{1}{6}$. Observamos que salió un 2, la información obtenida es:

$$\mathsf{I}(A = 2) = -\log_2\left(\frac{1}{6}\right) = 2.58 \text{ bit}$$

[2] Una variable aleatoria es justamente eso, una variable que tomará uno de varios posibles valores, cada uno con una determinada probabilidad previa

[3] En este ejemplo vemos que 1 bit equivale a la información ganada al observar un evento de probabilidad 0.5, o equivalentemente, una variable aleatoria binaria equiprobable. Los 0 y 1 de una computadora son 1 bit de información, porque toman esos dos estados, al igual que *Cara* y *Cruz*, y en computación se eligen códigos tal que ambos números ocurran aproximadamente igual cantidad de veces.

En los dos ejemplos previos se redujo completamente la incertidumbre con respecto a la variable aleatoria. Esto no tiene porqué ser necesariamente así, como ilustra el siguiente ejemplo:

(3) Tiramos un dado, pero el resultado lo ve otra persona y nos dice que el número obtenido es impar. Esto tenía una probabilidad $\mathsf{P}(A = Par) = 0.5$ de ocurrir, con lo que ganamos 1 bit de información. Sin embargo aún mantenemos incertidumbre con respecto al número exacto. Si luego la persona nos dice que salió un 5, esto tenía $\frac{1}{3}$ de probabilidad de ocurrir, dado que ya sabíamos que la el número era impar. La notación para esta probabilidad condicional es $\mathsf{P}(A = 3|A = Par) = \frac{1}{3}$. La información obtenida en esta segunda observación es de:

$$\mathsf{I}(A = 3) = -\log_2\left(\frac{1}{3}\right) = 1.58\ \mathsf{bit}$$

Notemos que la cantidad de información es equivalente cuando la adquirimo de a partes que cuando la adquirimos toda junta

Teorema de Bayes

El ejemplo 3 también ilustra el teorema de Bayes:

$$\mathsf{P}(X, Y) = \mathsf{P}(X|Y)\mathsf{P}(Y) \tag{1}$$

donde $\mathsf{P}(X, Y)$ es la probabilidad de que ocurran X e Y juntos, y $\mathsf{P}(X|Y)$ es la probabilidad de que ocurra X dado que ocurre Y. En el caso del ejemplo[a]:

$$\mathsf{P}(A = 3, A = impar) = \mathsf{P}(A = 3|A = impar)\mathsf{P}(A = impar) = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$$

[a] El ejemplo puede resultar un poco confuso porque siempre que el dado es 3 es impar, pero la relación condicional puede ser menos tajante. Es decir, en este caso particular $\mathsf{P}(A = 3, A = impar) = \mathsf{P}(A = 3)$, pero esto no siempre es así.

También es importante notar que la información obtenida al observar el valor que toma una variable aleatoria depende de qué valor se observa se oberva:

(4) Supongamos que tiramos una moneda sezgada, que sabemos que tiene una probabilidad de 0.99 de caer *Cara* y 0.01 de caer *Cruz*. Observamos el resultado de tirar la moneda, si cae *Cara* obtenemos $\mathsf{I}(A = Cara) = -log_2(0.99) = 0.0145\ \mathsf{bit}$ de información con la observación, pero si observamos *Cruz* obtenemos $\mathsf{I}(A = Cruz) = -log_2(0.01) = 3.32\ \mathsf{bit}$ de información.

De la definición formal de información junto con los ejemplos se pueden extraer algunas conclusiones. Por ejemplo, observar un evento es informativo en la medida en que el mismo sea inesperado. Observar un evento que teníamos alta confianza que ocurriría es poco informativo. De cierta forma, la información cuantifica la reducción de incertidumbre o la sorpresa que me genera una observación[4].

Pasamos a unos ejemplos más intuitivos de estos conceptos aplicados a la vida diaria:

(5) El tiempo de mañana es una variable aleatoria que puede tomar varios valores, los que podemos agrupar en $\{Llueve, Seco\}$. Despertarme mañana y observar el evento de si llueve o no, me aportará información.

(6) Si alguien me dice que hoy está lloviendo, y esto es algo que yo ya sabía, lo que me dijeron no me aportó información.

(7) Si el día de hoy me hace pensar que mañana lloverá, recibir un pronóstico de que lloverá mañana me aporta poca información. Por el contrario, recibir un pronóstico de que mañana no lloverá me aporta mucha información. Ganar mucha información está asociado a recibir una gran sorpresa, a que las expectativas que tenía no se cumplan.

(8) Si ayer no miré el valor del cierre del dólar, y hoy me entero que cerró a 31 pesos, no gano mucha información. Si por el contrario me entero que ayer el dolar cerró a 15 pesos, gané mucha información. En este caso, los posibles valores que puede tomar la variable aleatoria *Precio del dólar* son muchos, todos los valores mayores que 0, pero no considero que todos los valores tienen la misma probabilidad.

3 Fuentes de información, códigos y mensajes

En teoría de información se estudia cómo codificar y transmitir mensajes que contienen alguna información, por lo que ahora debemos definir qué es un mensaje y qué es un código. Comenzamos con una introducción semi formal y luego presentamos ejemplos intuitivos.

Una fuente de información es algún proceso en el mundo que puede tomar diferentes estados. Un código es una función o un conjunto de reglas que son utilizadas para producir un mensaje informando sobre los estados que toma la fuente de información. El código mapea los estados de la fuente de información a los valores que toma el mensaje. El mensaje luego es transmitido por un canal de transmisión. Finalmente es recibido en el otro extremo de canal por un destinatario, que si conoce el código utilizado puede decodificar el mensaje y ganar información sobre los estados que tomó la fuente de información.

Para que el mensaje sea informativo, el observador no debe saber su contenido de antemano. Es decir, desde el punto de vista del observador, antes

[4] Utilizamos las palabras observación y evento en un sentido semi matemático, más amplio que el sentido coloquial. Un evento es la realización de una de varias posibilidades, y una observación es obtener información sobre la ocurrencia de un evento. Por ejemplo, averiguar cual fuel el precio de cierre del dólar en la jornada de ayer es observar un evento (siendo el evento el precio del dólar).

de recibir el mensaje este puede tomar varios valores distintos, cada uno con una determinada probabilidad. Esta descripción que hacemos del mensaje es equivalente a la definición de variable aleatoria, es decir, el mensaje es una variable aleatoria a los ojos del receptor. A medida que el observador va leyendo el mensaje, va reduciendo su incertidumbre sobre el valor que toma el mensaje por lo que gana información. Es importante notar que no es necesario que el mensaje sea producido de forma aleatoria para que el mismo parezca una variable aleatoria para el observador. Más adelante se ejemplificaremos este último punto.

Hay mensajes de tipo continuo y discreto[5]. Los mensajes discretos consisten en una secuencia de símbolos. Los símbolos provienen del conjunto de símbolos disponibles en el código utilizado (este conjunto se llama el diccionario del código), y son elegidos de acuerdo a los valores que toma la fuente de información y las reglas del código.

Algunos ejemplos de de posibles diccionarios para escribir un mensaje discreto son:

(9) Un diccionario binario de 0 y 1, las letras del abecedario, las palabras del idioma español, las letras {A, T, G, C}, o un diccionario compuesto de las palabras {*Cara*, *Cruz*}[6].

Ahora que definimos qué es un mensaje discreto, veamos algunos ejemplos:

(10) Enviaremos un mensaje sobre los estados que toma una moneda que un mono está tirando repetidamente. La fuente de información en este caso es el mono tirando la moneda. Como código elegimos que cuando se obtiene *Cara*, escribimos un 0 en un papel, y cuando sale *Cruz* escribimos un 1. El mensaje consistirá entonces en una secuencia de 0s y 1s, y se verá aproximadamente algo así: $\{1, 1, 0, 0, 1, 1, 1, 1, 0, 1 \ldots\}$

En el ejemplo 10 la fuente de información produce el mensaje de manera completamente aleatoria, y sin ninguna estructura. Como en cada símbolo es igual de probable observar 0 que 1, al observar un símbolo del mensaje siempre ganamos 1 bit de información. Es decir, si observamos un mensaje de 10 símbolos de largo, habremos obtenido 10 bits de información[7].

[5] Discutiremos principalmente mensajes discretos porque son más fáciles de tratar matemática e intuitivamente, pero todo aplica también a mensajes continuos.

[6] Obviamente los nombres específicos de los elementos del diccionario no son relevantes, da lo mismo si les llamamos 0 y 1 o *Cara* y *Cruz*. Por esto, lo único relevante del diccionario es la cantidad de elemento que tiene.

[7] Esta es una propiedad matemática importante de la definición de información. Si tengo N eventos independientes que pueden ocurrir con probabilidades $p_1, p_2, \cdots p_N$, la probabilidad de que ocurran todos esos eventos es $p = p_1 \times p_2 \cdots \times p_N$. Como $\log(a \times b) = \log(a) + \log(b)$, la información que gano al observar esta combinación de eventos independientes es la suma de la información que obtengo de observar cada evento, lo que se ajusta a nuestra intuición. En el ejemplo del texto, observamos N eventos que son los N resultados de N tiradas de la moneda. La probabilidad de ver una determinada combinación de 3 resultdos es de la probabilidad de cada resultado, en este caso $0.5 \times 0.5 \times 0.5 = 0.5^3$. La definición de información me dice que

(11) Estamos leyendo en órden la lista de aprobadas del curso de neurociencias. Se puede decir que el curso junto con las evaluaciones y las estudiantes es la fuente de información. Codificamos el estado de la fuente de información escribiendo un 1 en un papel cuando la alumna aprueba y un 0 cuando desaprueba el curso. El mensaje se verá aproximadamente así: $\{1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0 \ldots\}$[8].

En en ejemplo 11 la fuente de información no sería en principio aleatoria, pero el mensaje lo parece. Sabiendo de antemano que hay un porcentaje de exoneración del 70%, cuando observamos un 1 obtenemos $\mathsf{I}(A = Aprobada) = -log_2\,(0.7) = 0.514$ bit, y cuando observamos un 0 obtenemos $\mathsf{I}(A = No\ Aprobada) = -log_2\,(0.3) = 1.73$ bit de información.

(12) Estamos leyendo una secuencia de ADN. La fuente de información es la molécula que estamos secuenciando. El secuenciador codifica la identidad de las bases en la secuencia utilizando el diccionario $\{A, T, C, G\}$. Un mensaje de este tipo tendría el siguiente aspecto: $\{\ldots T, T, A, C, G, A, \ldots\}$.

En el ejemplo 12 la fuente de información posiblemente tenga un alto nivel de estructura, definiendo estructuras biológicas importantes, pero aún así el mensaje parece aleatorio a nuestros ojos. Si no sabemos nada de la estructura de este mensaje, asumiendo que las 4 bases son igual de probables, cada símbolo observado del mensaje nos da $\mathsf{I}(A = Base_i) = -log_2\left(\frac{1}{4}\right) = 2$bit de información.

(13) Volviendo al caso del ADN, la molécula en sí es un mensaje que contiene información sobre las secuencias de aminoácidos de las proteínas. El código genético es un código de la misma forma que los que venimos analizando, que mapea de tripletes de bases (codones) a un diccionario compuesto por 21 símbolos: los 20 aminoácidos estándares más un elemento que señaliza el fin del mensaje. Este código es degenerado, ya que diferentes tripletes se asocian a diferentes aminoácidos. Es interesante en este caso preguntarse cuál sería la fuente de información.

Otra propiedad que vale la pena notar de los mensajes, es que no tienen porqué representar completamente lo que ocurre con la fuente de información. Es decir, luego de observar un mensaje puede seguir quedando incertidumbre de la fuente de información, como ilustra el siguiente ejemplo:

(14) Tomamos como la fuente de información las tiradas de un dado. Luego elaboramos un mensaje con el siguiente código: escribimos *Par* cuando el resultado es 2, 4 o 6, e *impar* cuando es 1, 3 o 5. Al leer el mensaje

la información obtenida al observar los 3 eventos considerados en conjunto es la misma que la obtenida al observarlos de a 1 (para que aplique esto es muy importante que los eventos sean independientes entre sí).

[8] La muestra se generó aleatoriamente utilizando el porcentaje de exoneración del curso pasado.

obtenemos información de la fuente de información, pero aún queda incertidumbre.

Hasta ahora el concepto y los ejemplos de mensajes que vimos quizás no dejan claro cómo los mismos pueden aplicarse a los sistemas sensoriales, pero son fácilmente extendibles para este fin. En el caso de los animales podemos pensar que el mundo es una fuente de información (es decir, una variable aleatoria), y que los sistemas sensoriales se encargan de recibir mensajes emitidos por este, que nos informan del estado del mundo.

(15) Pensemos que llego a un lugar que desconozco y veo una ventana. La estructura del mundo del otro lado de esa ventana es una variable aleatoria que desconozco. En este caso el mundo exterior es la fuente de información, pero es una variable aleatoria mucho más compleja que una moneda, un dado o una secuencia de símbolos, dado que tiene muchos más estados posibles que puede tomar. Por ejemplo, en cada posible evento hay distintos objetos, en distintas disposiciones, distintos colores, texturas, etc. Son muchos los posibles estados que puede tomar la fuente de información, por lo tanto tengo incertidumbre con respecto a la misma.
Los objetos del mundo reflejan luz que puede llegar a mis ojos. El patrón específico de luz que entra a mis ojos depende del estado del mundo exterior, y mi posición en el mismo. Este patrón de luz produce un determinado patrón de activación en los fotorreceptores de mi retina. Podemos considerar el patrón de activación de los fotorreceptores de mi retina como el mensaje que le llega a mi sistema nervioso[9]. Antes de mirar por la ventana son muchos los posibles patrones de activación posibles que pueden darse en mis fotoreceptores, cada uno con una determinada probabilidad. Una vez que miro por la ventana se genera el patrón de activación en mi retina (el mensaje) de acuerdo con el estado del mundo exterior (la fuente de información) y un código que traduce de estados del mundo exterior a patrones de activación en mi retina. Si sé el código que relaciona los estados del mundo exterior a los patrones de activación en la retina, puedo decodificar el estado del mundo exterior[10].
También, según mi conocimiento previo de lo que hay fuera de la ventana, el mensaje que reciba será más o menos informativo. Si pienso que miraré por la ventana para ver una pared de ladrillos, efectivamente ver una pared de ladrillos me resultará poco informativo. Si en cambio el patrón de activación en mi retina es incompatible con una pared de ladrillos, será un mensaje muy informativo.

En el ejemplo 15 vemos que podemos aplicar el concepto de fuente de información y mensaje a elementos y procesos más complejos que los que habíamos

[9]Debemos ser inclusivos con las especies que no tienen cerebro pero sí procesan información visual, como los insectos.

[10]Este código muy complejo fue hallado por la evolución, ya que nuestro sistema nervioso rutinariamente decodifica el estado del mundo exterior a partir de este tipo de mensajes, pero aún no ha sido completamente encontrado por la ciencia, como ilustra el hecho de que aún no hay robots que decodifiquen fielmente el estado del mundo exterior a partir de una cámara de video

analizado. Los casos que ya vimos de mensajes eran unidimensionales, con una secuencia de valores ocurriendo en el tiempo. En el ejemplo 15 es difícil pensar en la dimensionalidad de la fuente de información. La dimensionalidad del mensaje quizás sea un poco más fácil: cada fotorreceptor tomará un valor de activación, por lo tanto tenemos tantas dimensiones como fotorreceptores en los dos ojos, más la dimensión del tiempo.

4 Entropía y redundancia

Habiendo introducido los conceptos de información, fuente de información, mensaje y código, pasamos a presentar los conceptos de entropía y redundancia.

La entropía de una variable aleatoria se define como el ritmo **promedio** al que la misma 'produce' información, o de modo similar, la cantidad de incertidumbre que tenemos sobre la misma (recordemos que cuanto mayor incertidumbre tenmos sobre una variable, más información ganamos al observarla). Otra forma de definirla es como el valor esperado de información que obtendremos al observar dicha variable aleatoria.

Matemáticamente, el valor esperado o promedio de una función de una variable aleatoria es la sumatoria de los valores de esa función, ponderada por la probabilidad de cada valor de la variable aleatoria. En este caso la función de la que queremos hallar el valor esperado es la información, por lo tanto la entropía de una variable aleatoria A que puede tomar valores $\{a_1, a_2, \ldots a_i, \ldots a_N\}$ con probabilidades $\{\mathsf{P}(a_1), \mathsf{P}(a_2) \ldots \mathsf{P}(a_N)\}$ es:

$$\mathsf{H}(A) = \sum_{i=1}^{i=N} -\mathsf{P}(a_i) \log\left(\mathsf{P}(a_i)\right) \tag{2}$$

donde $H(A)$ es la entropía de la variable aleatoria A.

Algunos ejemplos de cálculos de entropía:

(16) La variable aleatoria es una moneda con probabilidades $\mathsf{P}(Cara) = 0.5$ y $\mathsf{P}(Cruz) = 0.5$, con lo que la entropía es:

$$\begin{aligned} \mathsf{H}(moneda) &= -\mathsf{P}(Cara)\log\left(\mathsf{P}(Cara)\right) - \mathsf{P}(Cruz)\log\left(\mathsf{P}(Cruz)\right) \\ &= -0.5\log 0.5 - 0.5\log 0.5 = 1 \ \mathsf{bit} \end{aligned} \tag{3}$$

(17) Para un dado normal, tenemos 6 posibles eventos, cada uno con probabilidad $\mathsf{P}(dado = i) = \frac{1}{6}$, con lo que la entropía del dado es

$$\mathsf{H}(dado) = \sum_{i=1}^{i=6} -\mathsf{P}(dado = i)\log\left(\mathsf{P}(dado = i)\right) = \sum_{i=1}^{i=6} -\frac{1}{6}\log\frac{1}{6} = 2.58\mathsf{bit} \tag{4}$$

(18) Una moneda sezgada con probabilidad del 90% de obtener una cara y 10% de obtener cruz tiene una entropía de

$$\begin{aligned}\mathsf{H}(moneda) &= -\mathsf{P}(Cara)\log(\mathsf{P}(Cara)) - \mathsf{P}(Cruz)\log(\mathsf{P}(Cruz)) \\ &= -0.9\log 0.9 - 0.1\log 0.1 = 0.469 \text{ bit}\end{aligned} \quad (5)$$

De los ejemplos notamos que observar los resultados de una moneda sezgada en promedio nos da menos información que observar los de una moneda normal (en el extremo de una moneda que siempre caiga del mismo lado la información obtenida al observarla, al igual que su entropía, es 0). En general cuanto más equiprobables sean los diferentes estados de la variable aleatoria, mayor será su entropía.

Es importante mencionar que tanto las fuentes de información como los mensajes tienen entropía (ya que los dos son variables aleatorias), y que un mensaje y su fuente de información pueden tener entropías distintas. Ejemplos:

(19) Volvemos al ejemplo del dado y el código que produce un mensaje indicando si los resultados son *Par* o *Impar*. El dado, como ya calculamos, tiene una entropía de 2.58 bit. El mensaje producido toma dos valores diferentes de forma equiprobable, lo que, al igual que la moneda balanceada, le da una entropía de 1 bit.

(20) Introduciendo la noción de ruido (fundamental en teoría de la información, pero que aquí mencionamos brevemente), supongamos como fuente de información una moneda, y que se genera un mensaje en el que se transmiten los resultados usando números enteros, pero que hay ruido en este proceso (ya sea en la codificación o en la transmisión). Cuando sale *Cara* se codifica con un 1, y cuando sale *Cruz* se codifica con un 2, pero el ruido en el proceso hace que al número elegido se le pueda sumar o restar 1, ambos eventos con una probabilidad de $\frac{1}{3}$. Entonces el mensaje puede tomar los valoress $0, 1, 2, 3$ con probabilidades $\{\mathsf{P}(M = 0) = \frac{1}{6}, \mathsf{P}(M = 1) = \frac{1}{3}, \mathsf{P}(M = 2) = \frac{1}{3}, \mathsf{P}(M = 3) = \frac{1}{6}\}$. Si hacemos el cálculo, la entropía de la fuente de información es 1 bit, pero la entropía del mensaje es de 1.92 bit. Notemos también que ahora no siempre sabemos el resultado de la moneda luego de ver el mensaje.

Los ejemplos recien vistos son de mensajes obtenido de forma que cada símbolo es independiente de los anteriores (también se puede decir que son variables aleatorias sin memoria), pero este no es el caso en muchos mensajes interesantes. Un ejemplo de un mensaje con fuertes dependencias entre sus símbolos nos permite introducir el concepto de redundancia:

(21) Tomamos como variable aleatoria el libro 100 años de soledad, que genera una secuencia de letras extraídas del abecedario. Este es el primer enunciado de la novela, en el que cambiamos algunas letras por *.

> Muchos añ*s después, frente al pelotón de fusilamiento, el sargento Aureliano Buendía había de recordar aquella tarde remota en *** su padre lo llevó a conocer el *****.

> Es claro que la primera letra omitida es *o*, que las siguientes 3 son la palabra *que*, pero no es predecible que las últimas 5 son la palabra *hielo*.

El hecho de que podamos adivinar algunas de las letras ilustra el concepto de redundancia. Un mensaje es redundante cuando algunos de sus símbolos no ofrecen mucha información. En este caso, ya podemos tener certeza de algunas de las letras debido a que el mensaje tiene una fuerte estructura que conocemos: el lenguaje español, y esto hace que algunas partes del mensaje sean fácilmente predecibles a partir del resto del mismo.

Cuanto mayor certeza tenemos de cual puede ser un símbolo en un mensaje, mayor redundancia tiene el mismo. Por otro lado recordamos que cuanto mayor certeza tenemos de una variable aleatoria, menor es su entropía. Así vemos que los conceptos de entropía y redundancia están muy vinculados.

De cierta forma, la redundancia implica la presencia de estructura en un mensaje. Las palabras en un texto pueden ser predecibles el mismo tiene una estructura predecible, ya que los textos en general no son producidos mediante la selección aleatoria de palabras, sino que una palabra se elige en función de las que le rodean. Esto es importante, ya que en el mundo natural la información es altamente redundante, como veremos en breve. A su vez, podemos ver que la presencia de estructura reduce la entropía que podría llegar a tener un mensaje, ya que de cierta forma la estructura impone limitantes en los valores que puede tomar el mismo.

De forma más formal, la redundancia de un mensaje o código es la diferencia entre la entropía del mismo y la entropía máxima que permite el diccionario del código del mensaje. Por ejemplo, el mensaje generado por la moneda con 90% de chances de obtenre *Cara* tiene una entropía de $0.469\mathsf{bit}$, con lo que el código tiene una redundancia de $0.531\mathsf{bit}$. Es decir, estamos desaprovechando en promedio $0.531\mathsf{bit}$ por símbolo del mensaje. Si observamos un mensaje proveniente de esta moneda veremos muchos 0 juntos y pocos 1, lo que a pesar de que una tirada no dependa de las otras, genera estructura y patrones en el mensaje que podemos minimizar cambiando el código que utilizamos para transmitir los resultados del juego[11][12].

De esta sección extraemos las siguientes tres conclusiones: 1) que la impredecibilidad de los mensajes es una medida de información relevante que podemos cuantificar con la teoría de la información, 2) que los mensajes pueden contener patrones y regularidades que en general provienen de la estructura presente en la fuente de información de la que provienen, y 3) que un mensaje es más eficiente desde el punto de vista de densidad de información cuanto menos estructura predecible contenga.

Uno de los grandes aportes de la teoría de la información ha sido ayudar a desarrollar códigos que permiten comprimir la información de un mensaje para

[11] No entraremos en detalles de como hacer esto, pero vale la pena menionar que un nuevo código involucraría enviar en cada símbolo información sobre varios resultados juntos.

[12] Otra definición conceptual de la entropía de una fuente de información es la cantidad de símbolos en promedio que se necesitan para representar ese mensaje en un nuevo código de 0 y 1.

una transmisión y un almacenado más eficiente (por ejemplo con los formatos de almacenar imágenes, como jpg, o archivos en general como el zip).

Con los conceptos de entropía y redundancia podemos ver el primero de los ejemplos de aplicación de teoría de la información al estudio de los sistemas sensoriales: la hipótesis de la codificación eficiente.

5 Codificación eficiente en sistemas sensoriales

Uno de los primeros aportes de la teoría de la información a la neurociencia fue el desarrollo de la hipótesis de la codificación eficiente, o de reducción de la redundancia. Pueden encontrarse ideas vinculadas a esta hipótesis previo al desarrollo de la teoría de la información, pero esta permitió expresar la hipótesis en términos concretos, y estudiar de forma más profunda sus implicancias. La hipótesis fue planteada en los términos de teoría de la información por Fred Attneave en 1954 y por Horace Barlow en 1961, y desde entonces ha sido una de las principales hipótesis guía en el estudio teórico de los sistemas sensoriales.

La hipótesis plantea que una de las principales funciones de los centros que procesan los mensajes sensoriales es la de reducir la redundancia en los mismos. Esto equivale a decir que estos centros codifican estos mensajes de forma más eficiente, para aprovechar al máximo sus recursos. A continuación un ejemplo de juguete:

(22) En la figura 1 se observa un ejemplo de imagen con una alta redundancia que podríamos recodificar de forma más eficiente. La imagen consiste de $32 \times 22 = 704$ píxeles que pueden tomar un valor de 0 o 1. Si quisieramos enviar un mensaje codificando esta imagen, una alternativa sería enviar el valor de cada píxel, utilizando 704 bits. Pero la imagen tiene una alta cantidad de estructura, lo que le da una alta cantidad de redundancia. Si sabemos que estas imágenes siempre se componen de cuadrados de tamaño variable, sería mucho más económico reportar la posición del vértice superior izquierdo y el tamaño de cada cuadrado. Para especificar cuál de las 32 posibles coordenadas de X tendrá el vértice se requieren 5 bits, para especificar la coordenada de Y de entre las 22 posibles se requieren 4.5 bits, y para especificar el tamaño se requieren también 4.5 bits si tomamos que el tamaño máximo es de 22 cuadrados. Esto da que para codificar esta imagen, enviando sólo esas descripciones de los dos cuadrados necesitamos solamente 28 bits. Si el receptor del mensaje conoce el código utilizado, puede reconstruír los 704 píxeles a partir de esos 28 bits.

En el ejemplo 22 es fundamental para poder compirmir el mensaje que conozcamos la estructura que tiene el tipo de imágenes que queremos enviar. En el ejemplo asumimos que las mismas continen sólo cuadrados de tamaño variable, lo que nos permitió elaborar un plan de código más económico. Supongamos que tenemos imágenes similares pero con una estructura un poco diferente,

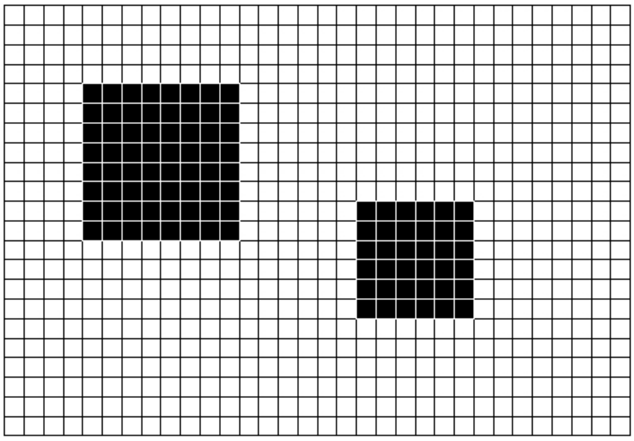


Figure 1: Imagen de 32x22 píxeles binarios, con un alto nivel de estructura.

pudiendo contener rectángulos y no sólo cuadrados. En este caso esta otra estructura requeriría otro código más caro, especificando el ancho y la altura del cada rectángulo.

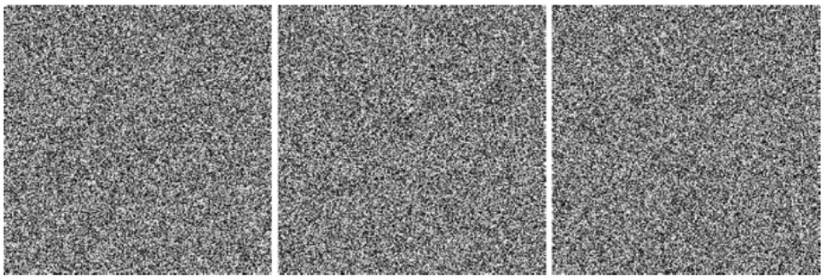
Esto último aplica a la codificación de cualquier tipo de fuente de información, para hacer un código eficiente es escencial comprender la estructura presente en los valores que toma la misma, que será la fuente de la redundancia en la señal.

La redundancia de los estímulos que reciben nuestros órganos sensoriales proviene de la estructura que existe en el mundo. Esta estructura puede apreciarse en la figura 2. Esta estructura proviene de las leyes de la física, y de la organización del mundo en objetos. Por ejemplo, los objetos del mundo tienen alguna extensión en el espacio, y son relativamente constantes a lo largo del tiempo, sin aparecer o desaparecer, y moviéndose a lo largo de trayectorias continuas. Esto hace que, por ejemplo, el mensaje sensorial que reciben mis ojos en un momento esté altamente correlacionado con el mensaje que estaba recibiendo instantes antes. De forma similar, el estado de activación de un receptor estará fuertemente correlacionado con el estado de activación de los receptores cercanos, ya que es probable que los mismos estén recibiendo información sobre el mismo objeto.

Esas correlaciones entre los receptores generan lo que llamamos **patrones** en la señal. En el caso del mundo natural, las imágenes generadas por el mismo tendrán un determinado conjunto de patrones recurrentes. Un código que contemple explícitamente estos patrones es lo que permite tener un código eficiente, ya que esto requiere un mensaje menor que si se estuvieran codificando todas las partes de dichos patrones cada vez que ocurren.

Entonces, en mayor detalle, la hipótesis de la codificación eficiente dice que los centros de procesamiento sensorial tienen la tarea de recodificar las señales que reciben para eliminar la redundancia generada por los patrones en dichas señales. Esto también tiene otras ventajas además de comprimir los mensajes

(a) Imagenes naturales, con la estructura característica de las mismas que es fuente de redundancia.



(b) Imágenes aleatorias, sin ninguna estructura ni rendundancia, en que el valor de cada píxel se elije de forma aleatoria.

Figure 2: Ejemplos de imágenes con y sin estructura etadística

sensoriales. Como se mencionó, los patrones en las señales sensoriales provienen de la estructura presente en el mundo. Esto significa que en general dichos patrones guardan información sobre los aspectos relevantes del mundo a nuestro alrededor. Entonces la capacidad de detectar un determinado patrón en la señal, además de permitir comprimir la misma, nos permite inferir sobre los eventos del mundo exterior.

(23) Un razgo fisiológico simple que se ajusta bien a este principio es la primera etapa de procesamiento de la información visual en la retina. El mensaje inicial que recibe el sistema visual son los valores de activación de los fotoreceptores en la retina (similar a los valores que pueden tener los píxeles en una imagen digital). Este mensaje no es muy útil en sí, ya que que no es explícita la información relevante del exterior. También es altamente redundante. En la primera etapa de procesamiento del sistema visual, el patrón de activación de los fotoreceptores es procesado por las células ganglionares de la retina. Las mismas tienen campos receptivos localizados, recibiendo conexiones de unos pocos fotoreceptores que son vecinos entre sí. Pero las células ganglionares no responden simplemente al valor de luz en esos receptores, sino que tienen un campo receptivo con una estructura centro-periferia antagónica. Es decir, la activación de los fotoreceptores en el centro del campo receptivo activa a la célula, y los fotoreceptores de la periferia la inactivan (o viceversa). Esto hace que ante un nivel de iluminación homogéneo en su campo receptivo, la célula no se active. De esta forma, las células ganglionares

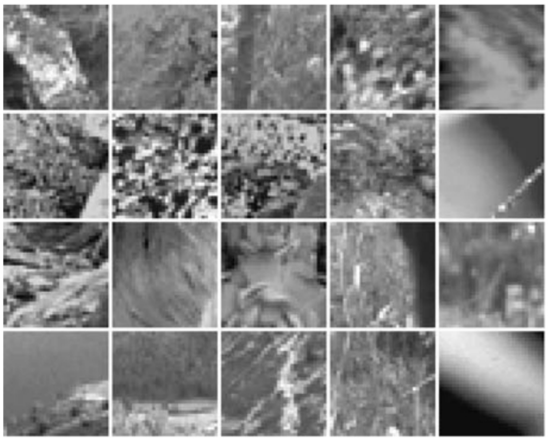
codifican los cambios de iluminación en la imagen que se forma en la retina, a diferencia de los fotoreceptores, que simplemente codifican el nivel de iluminación.

Esta característica de las células ganglionares puede relacionarse a una de las principales características de las imágenes: las mismas suelen contener áreas extensas de alta o baja iluminación, y las superficies generan zonas homogéneas de iluminación en la retina. Esto significa que en general los fotoreceptores vecinos tendrán activación similar, lo que genera una gran redundancia. El código empleado por las células ganglionares parece mucho más eficiente: señalar las zonas donde la iluminación cambia, y qué lado es el claro y cual el oscuro. Quien 'lea' el mensaje generado por las células ganglionares puede reconstruír la imagen original rellenando con iluminación homogénea las zonas de cambio señaladas por las mismas. Este cambio de código hace que haya mucho menos células ganglionares activadas que fotorreceptores activados, y que a su vez haya menor correlación entre la activación de las células ganglionares, haciendo que se pierda menos energía en codificar razgos redundantes de la señal.

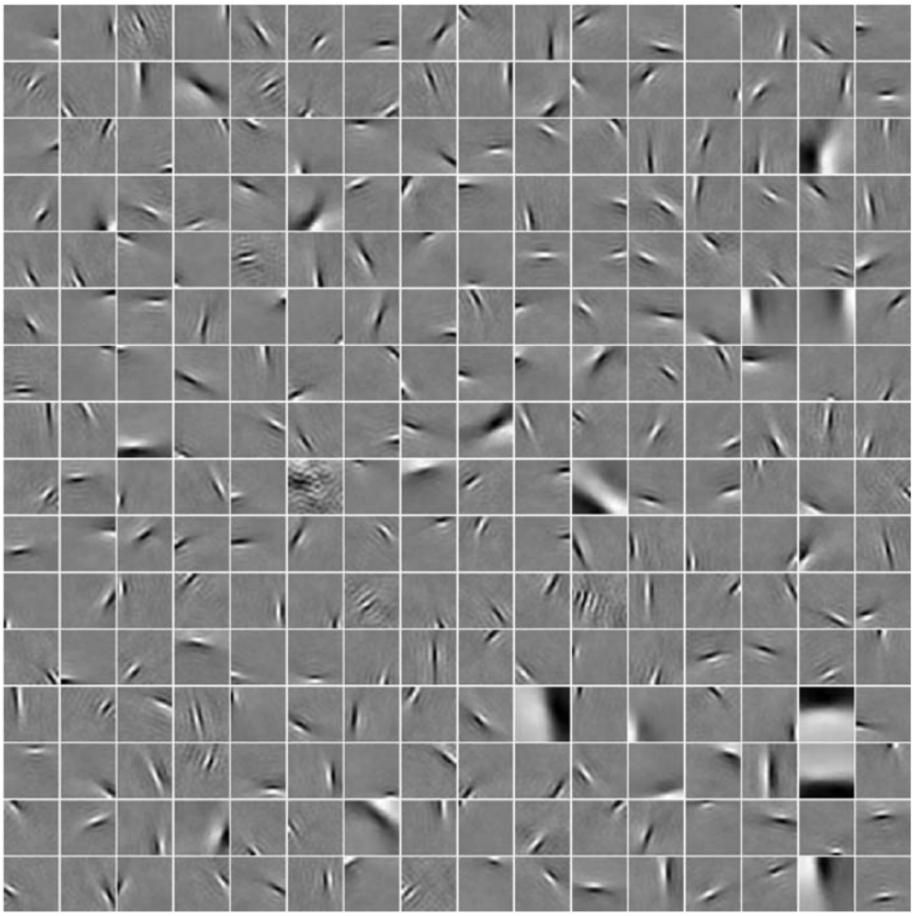
(24) Otro ejemplo interesante es la explicación del campo receptivo de las células de la corteza visual primaria, V1, en relación con estas ideas. Las células de la corteza visual primaria responden a estructuras orientadas, como los bordes y las rayas en las imágenes. Los campos receptivos de estas células tienen una estructura que hacen que sean activadas por este tipo de patrones. El *Independent Component Analysis* (ICA) es una técnica de análisis estadístico que, para el grupo de mensajes producido por una determinada fuente de información, intenta encontrar los patrones en dichos mensajes, de forma que los patrones encontrados sean independientes entre sí. Notemos que por la discusión sobre redundancia, representar patrones que sean independientes entre sí reduce la redundancia, ya que la presencia de uno no nos informa necesariamente de la presencia del otro.

El ICA puede aplicarse a una muestra de muchos miles de pequeños recortes de imágenes fotográficas (generalmente se usan recortes de aproximadamente 32 x 32 píxeles), como se muestran en la figura 3a. En este caso la fuente de informacion es el mundo y los mensajes son esos parches de imagen obtenidos. Cuando se aplica el ICA a las imágenes naturales, el mismo encuentra que los patrones que permiten reconstruír estas imágenes, y que tienen máxima independencia entre sí (bajo ciertos criterio), son estructuras orientadas como a las que responden las células de V1 (figura 3b. Esto significa que los campos receptivos ideales a los que se llega buscando independencia estadística, en relación con tener un código eficiente, son similares a los que están presentes en la corteza V1. Esta coincidencia debe tomarse con cautela, pero indica que es posible que el desarrollo de estos campos receptivos se de en parte por su característica de tener baja dependencias estadísticas, minimizando la redundancia.

En esta sección vimos cómo la teoría de la información permite expresar en



(a) Ejemplos de parches de imagen utilizados para encontrar los patrones regulares mediante ICA



(b) Campos receptivos encontrados mediante la aplicación del ICA a las imágenes naturales

Figure 3: Resultado de buscar los patrones presentes en las imágenes naturales mediante el método de ICA, maximizando la independencia entre dichos patrones

términos precisos los posibles procesos que pueden estar realizando los centros sensoriales, específicamente la hipótesis de la codificación eficiente. También vimos que, además de estudiar los sistemas biológicos, es importante estudiar el entorno que los rodea. Entender la estructura de las señales sensoriales que reciben los organismos a través de sus órganos sensoriales puede permitir comprender la estructura de estos. En este caso vimos cómo un estudio matemático de la estructura en las imágenes naturales muestra que los patrones codificados por las células de V1 son similares a los que se encuentran buscando patrones que sean estadísticamente independientes entre sí.

Es importante finalizar esta sección mencionando que la teoría de la información fue desarrollada para analizar sistemas muy distintos al sistema nervioso. Incluso el creador de la teoría de la información advirtió contra su aplicación por fuera de los sistemas que fue diseñado para analizar. Es un tema de debate actibo la aplicabilidad de la teoría de la información en su estado actual a la comprensión del sistema nervioso. El mismo puede ofrecer buenas intuiciones y conceptos, pero su aplicación debe hacerse con cuidado.

6 Referencias

1. *Possible principles underlying the transformations of sensory messages*, 1961, H. B. Barlow

2. *Some informtional aspects of visual perception*, 1954, F. Attneave

3. *Information theory and neural coding*, 1999, A. Borst, F. E. Theunissen.

4. *Spikes: Exploring the neural code*, 1997, F. Rieke, D. Warland, R. de Ruyter van Steveninck, W. Bialek.

5. *Natural image statistics: A probabilistic approach to early computational vision*, 2009, A. Hyvarinen, J. Hurri, P. O. Hoyer.