Teoría de la información

Presentación general

Año 2024

Núcleo de teoría de la información

- María Simon msimon@fing.edu.uy
- Alvaro Martín almartin@...
- Federico Lecumberry fefo@...
- Ignacio Ramírez nacho@...
- Máximo Pirri mpirri@...

Libros

- Thomas M. Cover, Joy A. Thomas: Elements of Information Theory, Ed. Wiley. Enfoque axiomático.
- Norman Abramson: Teoría de la Información y la Codificación, Ed. Paraninfo. Introducción intuitiva.
- Robert B. Ash: Information Theory, Dover Publications.
- Los propios artículos de Shannon en Bell Systems Journal o reeditados.

La información?

- Transmitir en el espacio
- Almacenar en el tiempo
- pero también información genética, procesamiento, percepción...

Cronología de la transmisión de la información y de la electricidad: *mucho antes que la teoría*

- 182X electricidad y su velocidad
- 1832 Morse *ferrocarril*
- 1860 Maxwell *Teoría de campos*
- Armstrong, Marconi *meteorología*
- máquina de Tesla
- 1890 teléfono
- 1900 radio AM
- 1905 Einstein
- 1914 18 Primera guerra mundial
- 1923 TV
- 1936 FM
- 1939 45 Segunda guerra mundial
- 1940 Radar, computadoras, criptografía, Turing.
- 1948 Shannon, *A mathematical theory of communications, Bell Systems Journal*

Después y ahora

- comunicaciones ópticas
- Orthogonal division multiplexing (ej. Televisión digital)
- Uso compartido de los medios (celulares, redes de computadoras)
- códigos al borde de la capacidad de los canales
- comunicaciones interplanetarias

Definición del problema y modelo

Claude E. Shannon.

30 de abril de 1916, Gaylord, Michigan, EEUU; 24 de febrero de 2001, Medford, Massachusetts, EEUU

1948: A mathematical theory of communication. Bell Systems Journal.

The fundamental problem of communication is that of reproducing at one point, either exactly or approximately, a message selected at another point.

Definición del problema y modelo

Claude E. Shannon.

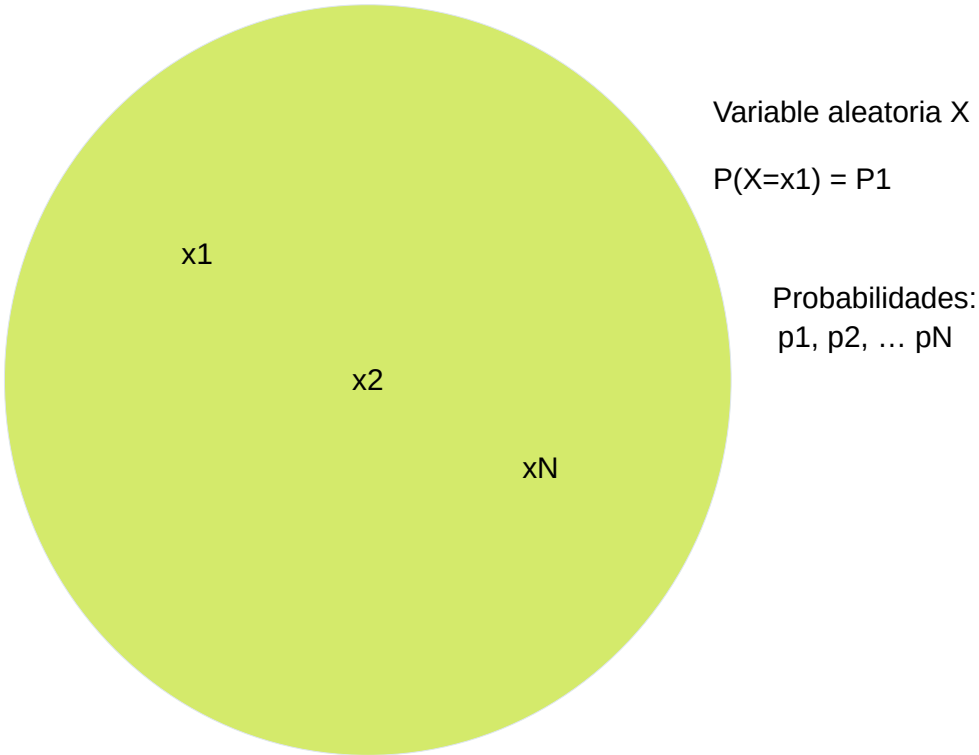
30 de abril de 1916, Gaylord, Michigan, EEUU; 24 de febrero de 2001, Medford, Massachusetts, EEUU

1948: A mathematical theory of communication. Bell Systems Journal.

The fundamental problem of communication is that of reproducing at one point, either exactly or approximately, a message selected at another point.

El problema fundamental de las telecomunicaciones es reproducir, en un punto, exacta o aproximadamente, un mensaje seleccionado en otro punto. Frecuentemente los mensajes tienen sentido, esto es: están relacionados en un cierto sistema de entidades físicas o conceptuales. Estos aspectos semánticos no se van a tener en cuenta en la descripción. El aspecto significativo es que el mensaje actual es seleccionado de un conjunto de mensajes posibles.

Fuente de mensajes finita y sin memoria



Es una variable aleatoria, en este caso finita.

Fuentes con memoria

- los lenguajes
- las imágenes
- cómo se puede aproximar a sin memoria.

Modelo de la información

«*Everything should be made as simple as possible, but not simpler*»(Einstein)

«*Entia non sunt multiplicanda praeter necessitatem*»(W. of Occam)

Modelo de la información

«*Everything should be made as simple as possible, but not simpler*»(Einstein)

«*Entia non sunt multiplicanda praeter necessitatem*»(W. of Occam)

Información asociada a un evento, como función de su probabilidad: $f : [0, 1] \longrightarrow R^+$

$I(x_i) = f(p_i)$ $\qquad$ $I(x_i) = f(p(X = x_i))$

Modelo de la información

«*Everything should be made as simple as possible, but not simpler*»(Einstein)

«*Entia non sunt multiplicanda praeter necessitatem*»(W. of Occam)

Información asociada a un evento, como función de su probabilidad: $f : [0, 1] \longrightarrow R^+$

$$I(x_i) = f(p_i) \qquad I(x_i) = f(p(X = x_i))$$

$$p_i > p_j \Longrightarrow f(p_i) < f(p_j)$$

Modelo de la información

«*Everything should be made as simple as possible, but not simpler*»(Einstein)

«*Entia non sunt multiplicanda praeter necessitatem*»(W. of Occam)

Información asociada a un evento, como función de su probabilidad: $f : [0, 1] \longrightarrow R^+$

$$I(x_i) = f(p_i) \qquad I(x_i) = f(p(X = x_i))$$

$$p_i > p_j \Longrightarrow f(p_i) < f(p_j)$$

$$p_i = 1 \Longrightarrow f(p_i) = 0$$

Modelo de la información

«*Everything should be made as simple as possible, but not simpler*»(Einstein)

«*Entia non sunt multiplicanda praeter necessitatem*»(W. of Occam)

Información asociada a un evento, como función de su probabilidad: $f : [0, 1] \longrightarrow R^+$

$$I(x_i) = f(p_i) \qquad I(x_i) = f(p(X = x_i))$$

$$p_i > p_j \Longrightarrow f(p_i) < f(p_j)$$

$$p_i = 1 \Longrightarrow f(p_i) = 0$$

$$p_i = 0 \Longrightarrow f(p_i) \to \infty$$

Modelo de la información

«*Everything should be made as simple as possible, but not simpler*»(Einstein)

«*Entia non sunt multiplicanda praeter necessitatem*»(W. of Occam)

Información asociada a un evento, como función de su probabilidad: $f : [0, 1] \longrightarrow R^+$

$$I(x_i) = f(p_i) \qquad I(x_i) = f(p(X = x_i))$$

$$p_i > p_j \Longrightarrow f(p_i) < f(p_j)$$

$$p_i = 1 \Longrightarrow f(p_i) = 0$$

$$p_i = 0 \Longrightarrow f(p_i) \to \infty$$

$$I(x_i, x_j) = I(x_i) + I(x_j) \qquad f(p_i p_j) = f(p_i) + f(p_j)$$

Entonces f es... la función logaritmo

$$f(p) = A\log(\tfrac{1}{p}) = -A\log p$$

Algunas propiedades:

- $\log(ab) = \log a + \log b$
- $\log(\frac{1}{a}) = -\log a$
- $\log_a c = \log_a(b)\log_b(c)$
- La constante A se absorbe en la base del logaritmo.

Valor esperado y entropía

El valor esperado de $f(X)$ en la probabilidad p es
$E_p(f(X)) = \sum p(X = x_i) f(x_i)$

Valor esperado de la información:
$H(X) = E_p(I)$
$H(X) = -\sum p(X = x_i) \log p(x_i) = -\sum p(x) \log p(x)$

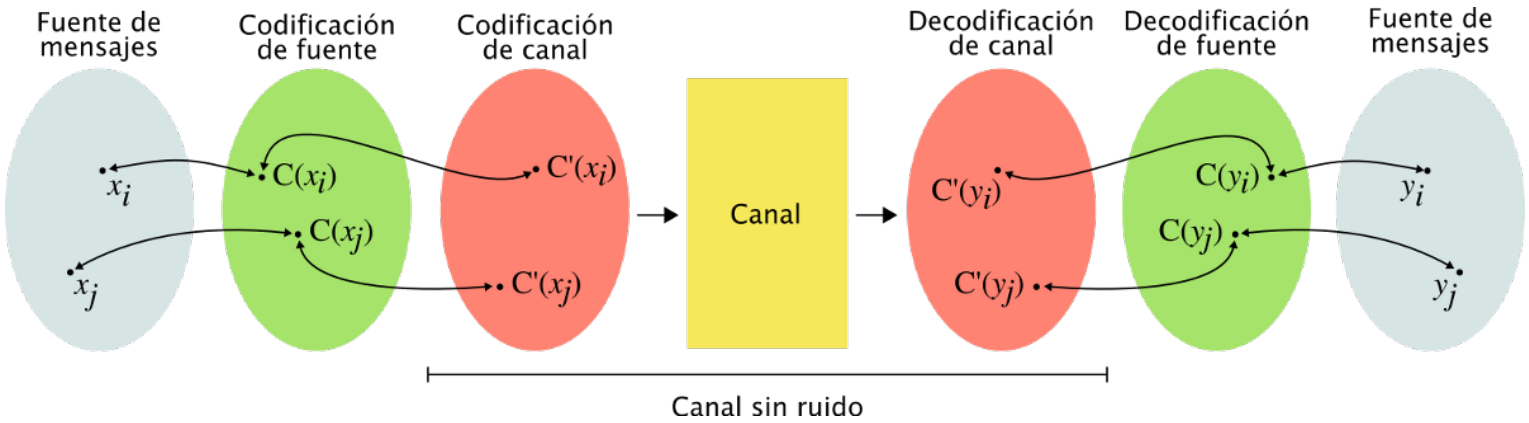
Componentes de un sistema

- Fuente de mensajes: discreta, finita, con o sin memoria. Se puede extender de discreta a numerable. Se tratarán en el curso también fuentes continuas.
- Codificación de fuente
- Alfabeto código: letras con las que se forman palabras
- Codificación de canal
- Canal: ruido, interferencia, distorsión, limitación de potencia, limitación de tamaño de archivos, tiempo o ancho de banda.

Esquema general

Codificación de fuente

Fuente generadora de mensajes de un alfabeto fuente $\mathcal{X} = \{x_1, \ldots, x_m\}$ con probabilidades $p_X(x_i)$. A cada uno de los mensajes se le asignará una palabra de código $C(x_i)$.



¿Cómo asignar las palabras de códigos de forma *óptima* y sistemática?

Dos teoremas: Primero

- LA FUENTE tiene una entropía
- Primer teorema o de la codificación sin ruido o de compresión: *El límite de la compresión sin pérdida es la entropía. No se puede comprimir más allá de la entropía, pero se puede acercarse indefinidamente a ella.*
- La demostración es constructiva

Dos teoremas: Primero

- LA FUENTE tiene una entropía
- Primer teorema o de la codificación sin ruido o de compresión: *El límite de la compresión sin pérdida es la entropía. No se puede comprimir más allá de la entropía, pero se puede acercarse indefinidamente a ella.*
- La demostración es constructiva

Un código tiene palabras de longitud l_i. Su longitud media es $L = E_p(l_i) = \sum p_i l_i$. $L \geq H(X)$ para cualquier código.

Se puede hallar códigos tales que $L \leq H(X) + \epsilon$

Primera parte: Compresión

El límite es la entropía, es alcanzable y el teorema es constructivo.
Largo medio: $L = E(l_i)$
$L \geq H(X)$
Ejemplo:

mensaje	prob	cod1	cod2	cod3	cod4
sol	$\frac{1}{2}$	00	0	1	0
nubes	$\frac{1}{4}$	01	10	10	10
lluvia	$\frac{1}{8}$	10	110	100	110
niebla	$\frac{1}{8}$	11	1110	1000	111

Aquí $H(X) = 1{,}75$ $L1 = 2$ y $L4 = 1{,}75$

Una idea precursora

MORSE CODE
(ALPHABETICAL)

A	● —	N	— ●
B	— ● ● ●	O	— — —
C	— ● — ●	P	● — — ●
D	— ● ●	Q	— — ● —
E	●	R	● — ●
F	● ● — ●	S	● ● ●
G	— — ●	T	—
H	● ● ● ●	U	● ● —
I	● ●	V	● ● ● —
J	● — — —	W	● — —
K	— ● —	X	— ● ● —
L	● — ● ●	Y	— ● — —
M	— —	Z	— — ● ●
1	● — — — —	6	— ● ● ● ●
2	● ● — — —	7	— — ● ● ●
3	● ● ● — —	8	— — — ● ●
4	● ● ● ● —	9	— — — — ●
5	● ● ● ● ●	0	— — — — —

Otros conceptos derivados directamente de la entropía

Entropía conjunta $H(X,Y)$
Entropía condicional o equivocación $H(Y|X)$
$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$
$H(X|Y) \leq H(X)$ Condicionar reduce la entropía.

Información mutua o correlación
$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$

Otros conceptos derivados directamente de la entropía

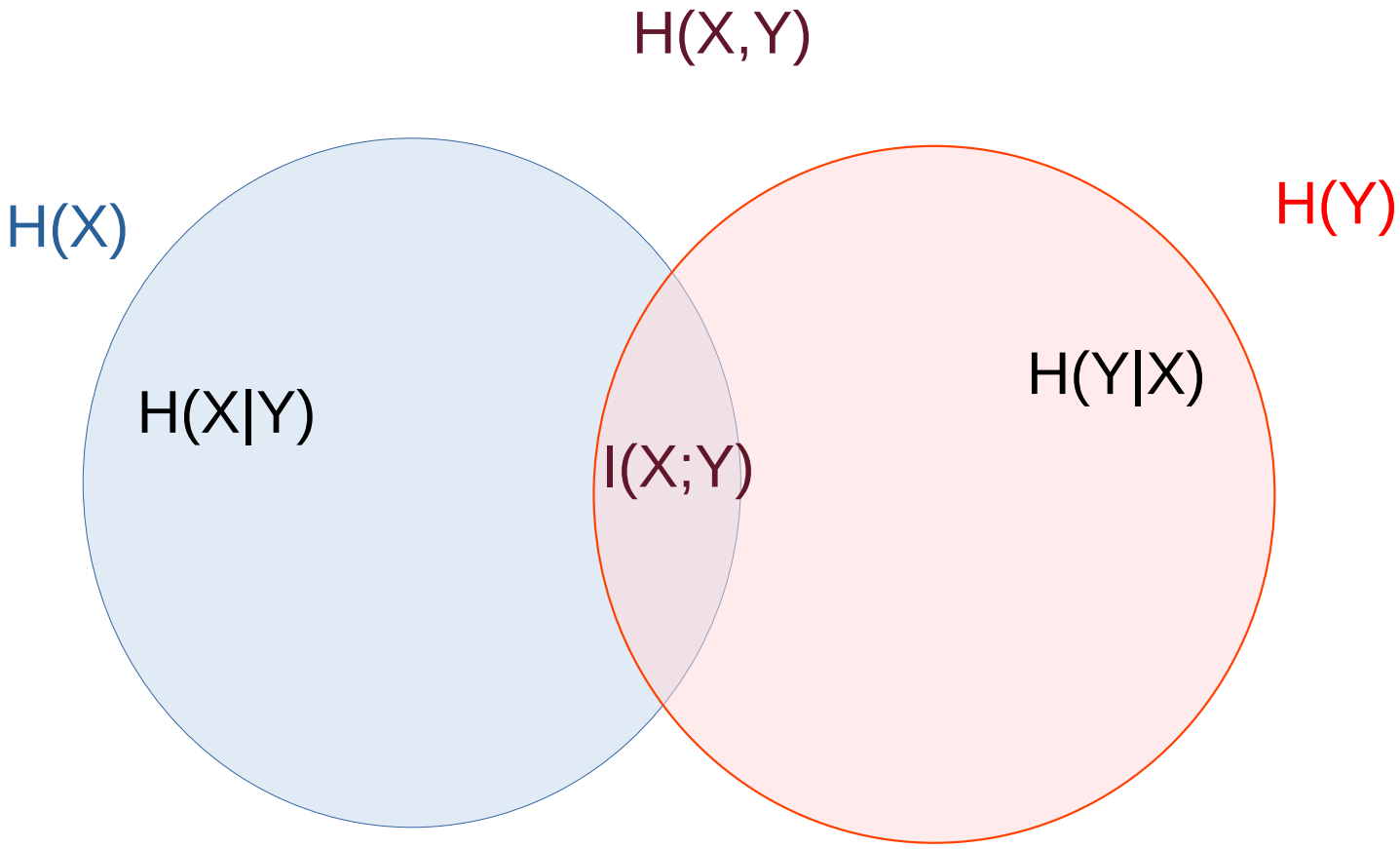
Entropía conjunta $H(X,Y)$
Entropía condicional o equivocación $H(Y|X)$
$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$
$H(X|Y) \leq H(X)$ Condicionar reduce la entropía.

Información mutua o correlación
$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$

El canal de transmisión se puede modelar como X, Y, aunque no es la única interpretación de la entropía conjunta. Un canal discreto, finito y sin memoria se puede representar mediante las probabilidades conjuntas o condicionales.
El máximo de la información mutua es la capacidad.

Recordar y usar la regla de la cadena:
$p(x,y) = p(x)p(y|x) = p(y)p(x|y)$

Diagrama representativo



$H(X,Y) <= H(X)+H(Y)$ $H(X,Y) = H(X)+H(Y|X) = \ldots$

$I(X;Y) = H(x)-H(X|Y) = H(Y)-H(Y|X)$

Dos teoremas: Segundo

- EL CANAL tiene una capacidad.
- Segundo teorema o de la codificación con ruido o de transmisión: *Se puede transmitir una cantidad de bits por símbolo menor que la capacidad, pero arbitrariamente cercana a ella, con probabilidad de error arbitrariamente baja*
- *El límite de la transmisión sin errores es la capacidad del canal*
- La demostración no es constructiva, pero es más que una prueba de existencia

Hay ruido inherente a la materia y ruido debido a otros usos del mismo medio.

Transmisión

El límite es la capacidad del canal, es alcanzable y el teorema es NO constructivo pero es MUCHO MÁS que una prueba de existencia.
Se renuncia a saber qué pasa con un código particular para sacar conclusiones sobre el promedio de los códigos posibles.

Transmisión

El límite es la capacidad del canal, es alcanzable y el teorema es NO constructivo pero es MUCHO MÁS que una prueba de existencia.
Se renuncia a saber qué pasa con un código particular para sacar conclusiones sobre el promedio de los códigos posibles.
En la práctica eso se logra introduciendo redundancia en forma controlada, para detectar o corregir errores. Ej: bit de paridad.

De lo digital a lo analógico

- La relación señal ruido o $\frac{P}{N}$ de un canal se vincula directamente a la probabilidad de error P_e
- El ancho de banda se vincula con la velocidad de señalización o tasa de símbolos por segundo

De lo digital a lo analógico

- La relación señal ruido o $\frac{P}{N}$ de un canal se vincula directamente a la probabilidad de error P_e
- El ancho de banda se vincula con la velocidad de señalización o tasa de símbolos por segundo

Capacidad de un canal en el mundo analógico: Fórmula de Hartley Shannon

$C = B_T log(1 + \frac{P}{N})$