WIKIPÉDIA

Teoria da informação

Origem: Wikipédia, a enciclopédia livre.

A **teoria matemática da informação** estuda a quantificação, armazenamento e comunicação da informação. Ela foi originalmente proposta por *Claude E. Shannon* em 1948 para achar os limites fundamentais no processamento de sinais e operações de comunicação como as de compressão de dados, em um artigo divisor de águas intitulado *"A Mathematical Theory of Communication"*. Agora essa teoria tem várias aplicações nas mais diversas áreas, incluindo *inferência estatística, processamento de linguagem natural, criptografia, neurociência computacional, evolução, computação quântica*, dentre outras.

A medida chave em teoria da informação é a **entropia**. A entropia é o grau de casualidade, de indeterminação que algo possui. Ela está ligada à quantidade de informação. Quanto maior a informação, maior a desordem, maior a entropia. Quanto menor a informação, menor a escolha, menor a entropia.[1] Dessa forma, esse processo quantifica a quantidade de incerteza envolvida no valor de uma variável aleatória ou na saída de um processo aleatório. Por exemplo, a saída de um cara ou coroa de uma moeda honesta (com duas saídas igualmente prováveis) fornece menos informação (menor entropia) do que especificar a saída da rolagem de um dado de seis faces (com seis saídas igualmente prováveis). Algumas outras medidas importantes em teoria da informação são **informação mútua**, **informação condicional** e **capacidade de um canal**

Aplicações de tópicos fundamentais da teoria da informação incluem compressão sem perdas de dados (e.g ZIP), e compressão de dados (e.g MP3 e JPEG).

O campo está na intersecção da matemática, estatística, ciência da computação, física, neurobiologia e engenharia elétrica. Seu impacto é crucial, por exemplo, no sucesso das missões da sonda *Voyager* no espaço, no entendimento desde buracos negros até da consciência humana como na teoria de integração da informação (do inglês, *Integrated Information Theory*), proposta por Giulio Tononi.

Índice

Contexto histórico

O marco que estabeleceu a teoria da informação e chamou imediatamente a atenção mundial foi o artigo *A Mathematical Theory of Communication* escrito por Claude Shannon de julho a outubro de 1948.

Antes deste artigo, algumas abordagens teóricas ainda que limitadas vinham sendo desenvolvidas nos laboratórios da Bell, todas implicitamente assumindo eventos de igual probabilidade. O artigo *Certain Factors Affecting Telegraph Speed* de Harry Nyquist escrito em 1924 contém uma seção teórica que quantifica *inteligência* e a velocidade de transmissão pela qual ela pode ser transmitida por um sistema de comunicação, estabelecendo a relação $W = K \log m$, onde W é a velocidade de transmissão da inteligência, m é o número de níveis de tensão para cada intervalo de tempo, e K é uma constante. Em 1928, Ralph Hartley publicou o artigo *Transmission of Information*, onde aparece a palavra *informação* como uma quantidade mensurável que a capacidade do destinatário distinguir diferentes sequências de símbolos, levando à expressão $H = \log S^n = n \log S$, onde S e n representam, respectivamente, o número de símbolos possíveis e o número de símbolos na transmissão. Inicialmente, a unidade natural da transmissão foi definida como sendo o dígito decimal, sendo, posteriormente, renomeada para hartley em uma clara homenagem. Alan Turing em 1940, durante a 2ª Guerra Mundial, aplicou ideias similares como parte da análise estatística para decifrar a criptografia da máquina alemã Enigma.

Boa parte da matemática por trás da teoria da informação com eventos de diferentes probabilidades foi desenvolvida para os campos da termodinâmica por Ludwig Boltzmann e J. Willard Gibbs. As conexões entre as entropias da informação e termodinâmica, incluindo as importantes contribuições de Rolf Landauer na década de 1960, são exploradas na Entropia termodinâmica e teoria da informação.

No artigo seminal de Shannon, introduz-se pela primeira vez um modelo quantitativo e qualitativo da comunicação, apresentando-a como um processo estatístico subjacente à teoria da informação. Shannon inicia seu artigo dizendo

"O problema fundamental da comunicação é reproduzir em um dado ponto, exata ou aproximadamente, uma mensagem produzida em outro ponto."

Com este artigo vieram à tona os conceitos

- de entropia da informação e redundância de uma fonte, e sua aplicação no teorema de codificação da fonte;
- de informação mútua e capacidade de um canal com ruído, incluindo a promessa de comunicação sem perdas estabelecida no teorema de codificação de canais-ruidosos;
- da lei de Shannon-Hartley para a capacidade de um canal Gaussiano;
- do bit - uma nova forma de enxergar a unidade fundamental da informação.

Variáveis aleatórias discretas

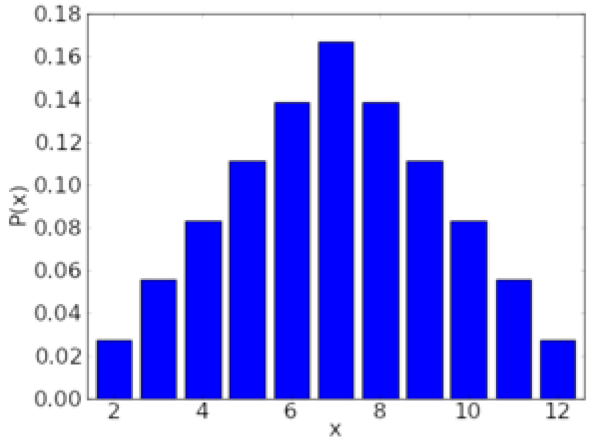
Antes de prosseguir é importante definir a notação utilizada para as variáveis aleatórias discretas. Dado uma variável aleatória R, que pode assumir m valores, podemos representa-la como:

$$R = \{r_1, \ldots, r_m\}$$

Onde r_i é o i-ésimo valor que pode ser assumido pela variável. Cada um dos m valores podem acontecer com probabilidade p, não necessariamente iguais. A distribuição de probabilidades de R é representada como:

$$p(R) = \{p_1, \ldots, p_m\}$$

Nesse caso p_i, com $1 \leq i \leq m$, representa a probabilidade do valor r_i acontecer. Esse tipo de distribuição pode ser representada com um gráfico de barras como na figura a seguir



Distribuição de probabilidades de uma variável discreta.

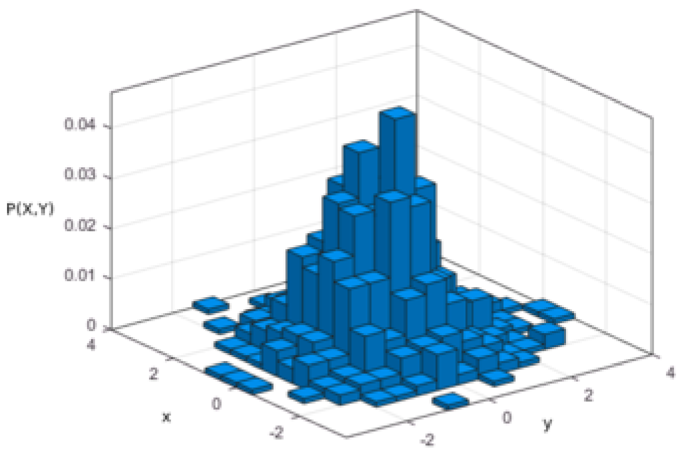
Para uma dada distribuição de probabilidades a condição $\sum_{i=0}^{n} p_i = 1$ é cumprida.

*

Dado duas variáveis aleatórias X e Y, cada uma podendo assumir quatro valores, logo $X = \{x_1, x_2, x_3, x_4\}$ e $Y = \{y_1, y_2, y_3, y_4\}$, com distribuições de probabilidade $p(X) = \{p(x_1), p(x_2), p(x_3), p(x_4)\}$ e $p(Y) = \{p(y_1), p(y_2), p(y_3), p(y_4)\}$. A distribuição conjunta das variáveis pode ser determinada medindo-se a frequência de ocorrência dos pares $B_{ij} = x_i, y_j$, com $i, j \leq 4$.

Medindo o número de ocorrências dos pares B_{ij}, em uma amostragem de tamanho n suficientemente grande, é possível determinar as probabilidades de cada par dividindo o número de ocorrências de cada um deles por n. Dando a distribuição de probabilidades conjunta (do inglês, *joint probability distribution*).

A distribuição conjunta pode ser por um histograma tridimensional, como o mostrado a seguir



Distribuição de probabilidade conjunta das variáveis X e .Y

À partir da distribuição conjunta de X, Y, é possível voltar as distribuições de X ou de Y (chamadas de distribuições marginais), através de um processo chamado *marginalização*. Por exemplo, se quisermos obter o valor da distribuição $p(X)$ para $X = x_i$ então devemos faze-lo somando por todos os m_y valores de Y para $X = x_i$:

$$p(X = x_i) = \sum_{j=0}^{m_y} p(x_i, y_j)$$

Fazendo isso fixando cada um dos possíveis valores de X, obtemos a distribuição marginal $p(X)$:

$$p(X) = \sum_{j=0}^{m_y} p(X, y_j) = \{p(x_1), p(x_2), p(x_3), p(x_4)\}$$

Para obter a distribuição marginal $p(Y)$ basta seguir o mesmo procedimento, fixando Y, em cada um de seus possíveis valores:

$$p(Y) = \sum_{i=0}^{m_x} p(x_i, Y) = \{p(y_1), p(y_2), p(y_3), p(y_4)\}$$

As grandezas da Teoria da Informação

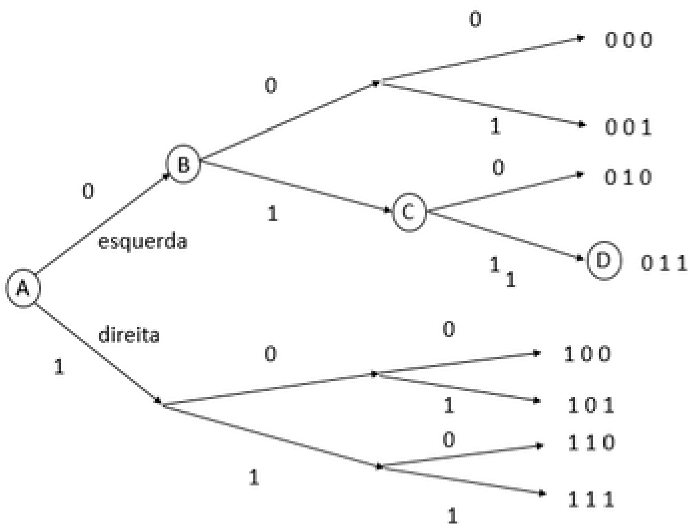
A teoria da informação é baseada na teoria de probabilidades e estatística. Ela preocupa-se com medidas de informação das distribuições associadas com variáveis aleatórias. Grandezas importantes da teoria da informação são a **entropia**, uma medida de informação de uma única variável aleatória, e **informação mútua**, uma medida de informação em comum entre das variáveis aleatórias.

Na formulação das grandezas da teoria da informação, escolheu-se utilizar bases logarítmicas (para manter propriedades como a aditividade da entropia), mais especificamente a entropia de Shannon é definida com logaritmos na base 2. A unidade utilizada para quantificar informação é o *bit* (baseado no logaritmo base 2), muito embora outras existam, como o *nat* (baseada no logaritmo natural), e o *hartley* (baseado no logaritmo na base 10).

Tendo citado as aplicações da teoria da informação, seu contexto histórico de surgimento e mencionado as grandezas envolvidas, chegou o momento de discutir mais a fundo alguns dos conceitos bases da teoria, começando pela unidade de informação, o *bit*.

Escolhendo destinos e o bit de informação

A fim de entender de forma intuitiva o que significa dizer 1 *bit* de informação, imagine a seguinte situação, um viajante, decide sair de sua cidade, no ponto marcado com a letra "**A**" na figura abaixo e chegar ao seu destino no ponto "**D**".



Caminho percorrido pelo viajante. Em cada bifurcação ele recebia a instrução esquerda (0) direita (1) até alcançar seu destino final.

O caminho entre "**A**" e "**D**", possui várias bifurcações (como os pontos "**B**" e "**C**"), como o viajante desconhece o caminho, em cada cidade que ele passa (representado pelas bifurcações) ele pede informação, perguntando se deve seguir a direita ou a esquerda. Na figura anterior dizer que ele deve seguir a esquerda é o mesmo que mostrar o *dígito binário* 0 a ele, e um sinal de que deve seguir a direita o mesmo que mostrar o dígito 1.

Dessa forma, como é possível ver pela figura, ele terá que pedir informação nos pontos "**A**", "**B**" e "**C**", note que independente do destino ($\{000, 001, 011, \dots\}$) o número de perguntas para alcança-lo (nesse caso) é sempre três. Ou seja, escolher entre oito destinos requer três perguntas:

$$8 \text{ destinos} = 2^3 \text{ destinos}$$

Note que o o expoente do número dois na equação anterior é igual ao número de perguntas feitas. Defini-se então que para escolher entre um dos oito possíveis destinos é necessária uma quantidade de informação igual a 3 bits.

Aplicando logaritmo de base dois na expressão anterior, temos:

$$3 = \log_2 8 \quad [\text{bits}]$$

É importante salientar, que como o viajante poderia ter escolhido qualquer um dos destinos finais possíveis, eles são todos equiprováveis, com probabilidade $p = 1/8$.

De forma análoga, para o caso em que se tem m possíveis destinos, e **supondo que o viajante possa escolher qualquer um deles com igual probabilidade**, a quantidade de informação, em bits, para alcançar um dos possíveis destinos é dada pela relação a seguir

$$n = \log_2 m \quad [\text{bits}]$$

Com isso conclui-se que n *bits* é a informação necessária para se escolher entre m alternativas equiprováveis, ou $1 bit$ é a quantidade de informação necessária para escolher entre duas alternativas equiprováveis.

$$n = \log_2 2 = 1 \text{ bits}$$

Informação de Shannon (h) ou surpresa

Usarei de outro exemplo para explicar a medida de Informação de Shannon h ou surpresa. Imagine nesse caso, uma moeda desonesta (enviesada), que tem probabilidade $p_{\text{cara}} = 0.9$ de dar cara e probabilidade $p_{\text{coroa}} = 0.1$ de dar coroa.

Por você estar acostumado que a jogada dessa moeda quase sempre dê **cara**, esse resultado não te surpreende. Mas um resultado **coroa** te surpreende por conta da "*raridade*" do evento. Pensando nisso, uma forma natural de se definir essa surpresa que se tem, seria como algo proporcional ao inverso da probabilidade p de ocorrência do evento, de modo que quando menor essa probabilidade maior a surpresa.

Shannon definiu essa grandeza como:

$$h = \log_2 \frac{1}{p}$$

Utilizando o logaritmo na base dois mantêm-se a propriedade de aditividade dessa grandeza.

Dessa maneira podemos calcular a surpresa da jogada da moeda retornar cara (h_{cara}) ou coroa (h_{coroa}) de acordo com a definição anterior.

$$h_{cara} = \log_2 \frac{1}{0.9} = 0.152$$

e,

$$h_{coroa} = \log_2 \frac{1}{0.1} = 3.322$$

De fato, $h_{coroa} > h_{cara}$, que surpresa!

Entropia (H)

A fim de chegar na formulação matemática da entropia, imagine por exemplo uma variável aleatória X, que pode assumir dois valores distintos x_1 e x_2 com probabilidades p_1 e p_2, respectivamente. Seguindo a notação definida na seção: Variáveis aleatórias discretas, temos:

$$X = \{x_1, x_2\}$$

e,

$$p(X) = \{p_1, p_2\}$$

A informação de Shannon associada a cada um dos valores é:

$$h_1 = \log_2 \frac{1}{p_1}$$

e,

$$h_2 = \log_2 \frac{1}{p_2}$$

Na prática, geralmente nós não estamos interessados em saber a surpresa de um valor em particular que uma variável aleatória pode assumir, e sim a surpresa associada com todos os possíveis valores que essa variável aleatória pode ter. De modo a obter a surpresa associada a todos possíveis valores que X pode assumir, defini-se a entropia $H(X)$ como a informação de Shannon média:

$$H(X) = p_1 h_1 + p_2 h_2 = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} = \sum_{i=1}^{2} p_i \log_2 \frac{1}{p_i}$$

Caso X, possa assumir m valores, a expressão anterior pode ser escrita de modo mais geral.

$$H(X) = -\sum_{i=0}^{m} p_i \log_2 p_i$$

A entropia do dado de seis faces

Uma aplicação direta para a equação da entropia definida anteriormente pode ser obtida com o exemplo um dado de seis faces. Representando o dado pela variável aleatória X, temos que:

$$X = \{1, 2, 3, 4, 5, 6\}$$

e,

$$p(X) = \{1/6, 1/6, 1/6, 1/6, 1/6, 1/6\}$$

Desse modo a entropia é:

$$H(X) = -\sum_{i=0}^{6} p_i \log_2 p_i = -6\frac{1}{6} \log_2 \frac{1}{6} = 2.585\,\text{bits}$$

Moeda boa, moeda má

Considere a moeda honesta, representada pela variável aleatória M_1:

$$M_1 = \{cara, coroa\} = \{0, 1\}$$

e,

$$p(M_1) = \{p_{cara} = 0.5, p_{coroa} = 0.5\}$$

E a moeda enviesada, como aquela utilizada para exemplificar a informação de Shannon, representada pela variável aleatória M_2.

$$M_2 = \{cara, coroa\} = \{0, 1\}$$

e,

$$p(M_2) = \{p_{cara} = 0.9, p_{coroa} = 0.1\}$$

Note que um resultado **cara** é representado pelo dígito 0 e um resultado **coroa** por um dígito 1. A entropia de cada uma das moedas pode ser calculada, sendo então:

$$H(M_1) = -\sum_{i=0}^{2} pilog_2 p_i = -2\frac{1}{2}\log_2\frac{1}{2} = 1bit$$

e,

$$H(M_2) = -\sum_{i=0}^{2} pilog_2 p_i = -(0.9\log_2 0.9 + 0.1\log_2 0.1) = 0.469bits$$

Nesse caso a entropia da moeda honesta é maior do que a da moeda desonesta, mas qual o significado disso? O que a medida de entropia pode me dizer? Essas questões serão abordadas na próxima seção, onde uma abordagem conceitual de entropia será tratada.

Uma visão conceitual da grandeza entropia

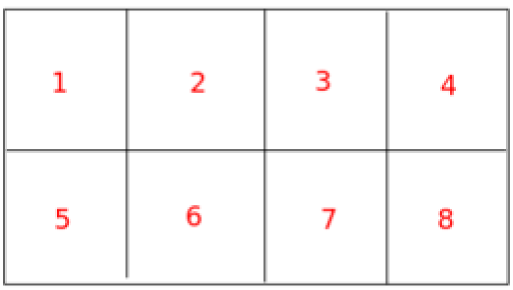
Fundamentalmente a entropia é uma medida de **incerteza**. Isso pode ser visto no exemplo anterior, na moeda honesta é muito difícil dizer qual será o resultado antes de joga-la, alguém pode arriscar dizer que o resultado será cara ou coroa, mas a *incerteza* continua maior do que no caso da moeda enviesada (logo com entropia também maior), onde podemos prever com certa tranquilidade que o resultado da jogada será cara.

No fundo, essa incerteza está ligada à previsibilidade do valor que será assumido por uma dada variável aleatória, prever um valor sorteado entre 100 valores equiprováveis (por exemplo, adivinhar o número que saíra em um dado de 100 faces) é mais difícil do que prever esse valor, caso esse dado seja enviesado e uma das faces tenha probabilidade alta de aparecer.

Agora sabemos também que a entropia é uma medida de informação, como uma coisa se relaciona a outra?

Justamente por ter mais incerteza sobre a possível saída de uma variável aleatória, você precisa de mais informação, para "*advinhar*" essa saída. Isso é análogo ao número de perguntas que o nosso viajante da seção "Escolhendo destinos e o bit de informação" teve de fazer para chegar ao seu destino. Desse modo quanto maior a entropia maior a incerteza e maior a informação que você precisa para "*advinhar*" uma possível saída que a variável aleatória pode apresentar

Para ilustrar o que foi dito, considere a seguinte situação, estacionaram seu carro em um estacionamento de 8 vagas dispostas como no desenho abaixo, para adivinhar em que vaga ele está você tem permissão de realizar perguntas sim ou não.



Esquema do estacionamento.

Você pode começar: "O carro está na direita?", caso a resposta seja sim, isso restringe as possíveis vagas pela metade, sendo elas as vagas 3, 4, 7 e 8. A próxima perguntar pode ser: "O carro está ao norte?", com uma resposta não, restam duas possibilidades, a vaga 7 ou 8. Uma ultima pergunta: "O carro está a direita?", basta para determinar em qual delas seu carro está. Nesse caso cada resposta sim/não te da 1 bit de informação. Como são necessárias três perguntamos podemos dizer que a entropia H é igual a 3 bits.

Assim fica fácil entender a ideia de que a entropia está relacionada com a quantidade de informação necessária para "*advinhar*" a resposta (ou uma possível saída de uma variável aleatória).

No exemplo das vagas, como a probabilidade do carro estar em qualquer uma das vagas é igual, cada bit de informação diminui o número de respostas possíveis pela metade.

Entropia da distribuição conjunta

A definição de entropia para uma distribuição conjunta $P(X, Y)$ pode ser obtida de forma direta da definição de entropia, por analogia, sendo:

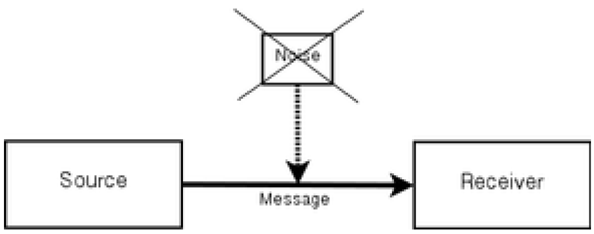
$$H(X, Y) = -\sum_{i=0}^{m_x} \sum_{j=0}^{m_y} p_{ij} \log_2 p_{ij}$$

Onde p_{ij} é a probabilidade de ocorrência do par $B_{ij} = x_i, y_j$. Essa definição é de particular importância para quando definirmos informação mútua.

Teorema de codificação da fonte

O teorema de codificação da fonte é fundamental para todos os meios de comunicação, uma vez que ele estabelece limites de como mensagens podem ser transmitidas e além disso, ele mostra que existem maneiras mais e menos eficientes de se fazer isso, dependendo da mensagem enviada. Aqui apenas uma ideia do que ele se trata será dada, para entendimento maior consulte as referências.

Antes de mais nada considere um canal, sem fonte de ruído, onde uma mensagem é codificada em sua fonte (*source*) por um *encoder*, enviada pelo canal até seu destino, decodificada por um *decoder* e interpretada pela pessoa alvo.



Esquema de um canal sem ruído.

Define-se a Capacidade do canal C como sendo numericamente igual ao número de dígitos binários comunicados por segundo. Se tivermos 1 bit por dígito binário a capacidade é definida em unidades de bits por segundo. Essa definição pode ser mais bem explorada matematicamente, mas para os fins aqui propostos, a definição dada é suficiente.

\par Dada as definições, imagine que se deseja transmitir uma série de m símbolos, representados pela variável $S = \{s_1, .., s_m\}$, sendo $p(S)$ a distribuição de probabilidades de S e $H(S)$ sua entropia. O teorema de codificação da fonte pode ser enunciado como segue:

"Dada a distribuição S com entropia $H(S)$, medida em bits por simbolo s, e um canal com capacidade C bits por segundo. Então é possível codificar os símbolos s enviados pela fonte de tal modo que a mensagem seja transmitida na capacidade máxima C do canal."

Enviando números de 1 a 8

Imagine uma fonte que envia números de 1 a 8, com igual probabilidade $p = 1/8$, então temos nossa variável $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$, podemos determinar a entropia de S como sendo $H(S) = \texttt{3bits/símbolo}$.

Caso os números sejam transmitidos por um canal com capacidade $C = 3bits/s$, o teorema de codificação da fonte garante que existe um modo de codificar os símbolos em S de modo tal que eles sejam transmitidos com capacidade máxima C, no caso 3 bits/s.

Um modo de codificar os valores de 1 a 8, seria representa-los por números binários de $3 = log_2 8$ dígitos binários, como na tabela a seguir, que indica o simbolo e sua respectivo código.

Simbolo	Código
1	000
2	001
3	010
4	011
5	100
6	101
7	110
8	111

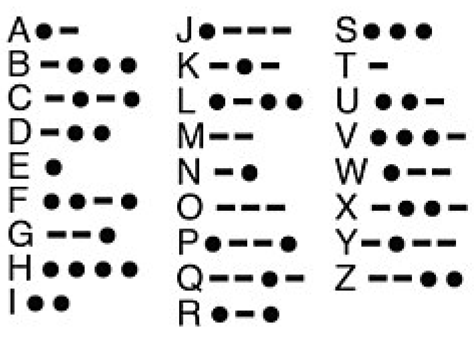
Sendo L o número de dígitos binários utilizado por código para cada simbolo de S. A eficiência ϵ é um número entre 0 e 1, calculada como a razão da entropia de S por L.

$$\epsilon = \frac{H}{L}$$

Nesse caso,

$$\epsilon = \frac{\texttt{3bits/símbolo}}{\texttt{3dígitosbinários/símbolo}} = \texttt{1bit/dígitosbinários}$$

Para esse caso simples é muito fácil encontrar a codificação necessária para transmitir os símbolos com máxima eficiência. Mas para maioria dos casos não é assim, e são necessários algoritmos mais rebuscados, como por exemplo a ***codificação de Huffman***, que não será discutida aqui, mas consiste em codificar os símbolos mais frequentes com códigos mais simples (que usam menos dígitos binários por exemplo). O código Morse (figura abaixo), se baseia nesse princípio, onde letras como o **E** mais frequentes na língua inglesa são representados por sequência mais simples, e outras letras menos frequentes como o **J** por sequências mais complicadas, isso ajuda a aumentar a eficiência com a qual a mensagem é enviada.



Código Morse.

É importante salientar que o código Morse precede o artigo de Shannon, sendo portanto desconhecidos esses limites teóricos para comunicar informação.

Informação Mútua

Considerações Iniciais

Dado duas variáveis aleatórias X e Y, a informação mútua $I(X,Y)$ entre elas é a quantidade de informação média que ganhamos sobre Y após observar um valor isolado de X

A informação mútua entre X e Y é definida como:

$$I(X,Y) = \sum_{i=1}^{m_x} \sum_{j=1}^{m_y} p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

Para m_x valores de X e m_y valores de Y. A expressão anterior pode ser trabalhada de forma a ser escrita como:

$$H(X,Y) = H(X) + H(Y) - I(X,Y)$$

Onde $H(X,Y)$ é entropia da distribuição conjunta dada já definida anteriormente.

Entropia condicional

Um modo alternativo de enxergar o conceito de informação mútua pode ser obtido considerando-se a entropia da saída em relação ao ruído do canal. Se não conhecemos o valor da entrada X então nossa incerteza sobre o valor de Y é dado por sua entropia $H(Y)$. Mas se conhecemos o valor de X então nossa incerteza sobre Y é reduzida de $H(Y)$ para um valor chamado de *entropia condicional* $H(Y|X)$, que é a incerteza média do valor de Y após X ser observado. Assim:

$$I(X,Y) = H(Y) - H(Y|X)$$

Como informação mútua e uma grandeza simétrica:

$$I(Y,X) = H(X) - H(X|Y)$$

Onde $H(X,Y)$ é a incerteza média que temos sobre o valor de X após temos observado Y, e portanto a incerteza média em X que não pode ser atribuída a Y.

Independência Estatística

Se X e Y são estatisticamente independentes, então conhecer um valor de X não nos dá nenhuma informação sobre Y e *vice versa*. Nesse caso cada valor de probabilidade da distribuição conjunta pode ser escrito como:

$$p(x_i, y_j) = p(x_i)p(y_j)$$

Substituindo na definição de informação mútua, e realizando algumas manipulações temos:

$$H(X) + H(Y) - H(X,Y) = 0$$

O que significa que $I(X,Y) = 0$, como esperado.

Entropia condicional e ruído

Diferente do que se considerou na seção onde tratamos do teorema de codificação da fonte, os canais geralmente possuem ruído (figura abaixo). Por esse motivo se considera que a saída do canal Y é igual a entrada X mas um ruído do canal η, é possível achar uma expressão do ruído do canal como a seguir

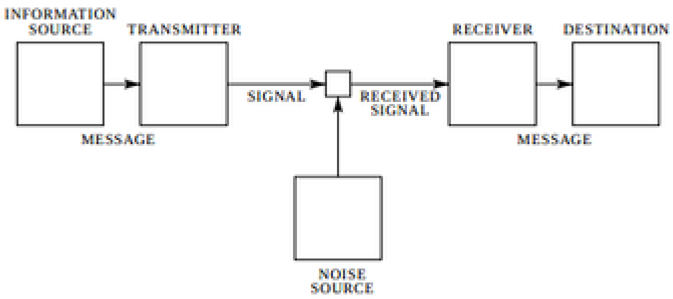


Fig. 1—Schematic diagram of a general communication system.

Esquema de canal com ruído.

$Y = X + \eta$

Da expressão $I(X,Y) = H(Y) - H(Y|X)$ nos leva a:

$I(X,Y) = H(Y) - H([X+\eta]|X)$

Se o valor de X é conhecido, então a incerteza em X é zero, logo ele não tem nenhuma contribuição na entropia condicional $H([X+\eta]|X)$, dando:

$I(X,Y) = H(Y) - H(\eta|X)$

Entretanto o valor do ruído η é independente do valor de X, logo $H(\eta|X) = H(\eta)$, o que nos permite reescrever a equação anterior como:

$I(X,Y) = H(Y) - H(\eta)$

Comparando as equações $I(X,Y) = H(Y) - H(Y|X)$ e a anterior podemos concluir que:

$H(Y|X) = H(\eta)$

Logo, a entropia do ruído é igual a entropia condicional $H(Y|X)$.

Referências

[1] Stone J. (2014). Information Theory: A Tutorial Introduction. Sheffield: Sebtel Press.

[2] MacKay D. (2003). Information Theory, Inference, and Learning Algorithms. Cambridge: Cambridge University Press.

[3] Shannon C., Weaver W. (1949). The Mathematical Theory of Communication. Urbana, IL: University of Illinois Press.

[4] Borst, A. \& Theunissen, F. Information theory and neural coding. Nature Neurosci. 2, 947–957

(1999).

[5] Tononi, 2012 Integrated information theory of consciousness: an updated account

Arch. Ital. Biol., 150 (2012), pp. 56–90.

Ver também

- Abraham Moles
- Distância Levenshtein
- Informação
- Norbert Wiener

Ligações externas

- "The Mathematical Theory of Communication", Claude Shannon
- Entropy on the World Wide Web
- Information Theory Resources

1. *WEAVER, W. A teoria matemática da comunicação. COHN, Gabriel (Org.). *Comunicação e indústria cultural*. 3. ed. São Paulo: Nacional, 1977.* [S.l.: s.n.]