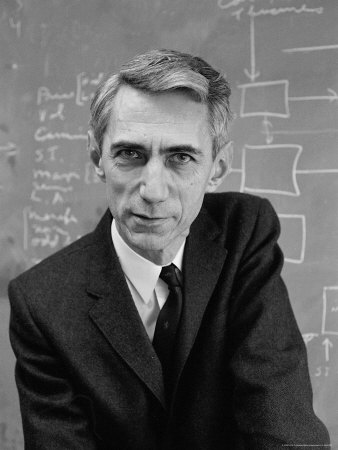
**Teoria da Informação**

**Entropia como medida de incerteza ou de informação:**



Claude Elwood Shannon (1916 — 2001) matemático e engenheiro eletrônico americano conhecido como "o pai da teoria da informação".

De 1932 a 1936, estudou matemática e engenharia elétrica na University of Michigan. Em 1948, publicou o importante artigo científico intitulado A Mathematical Theory of Communication enfocando o problema de qual é a melhor forma para codificar a informação que um emissor queira transmitir para um receptor. Neste artigo, trabalhando inclusive com as ferramentas teóricas utilizadas por Norbert Wiener, Claude Shannon propôs com sucesso uma medida de informação própria para medir incerteza sobre espaços desordenados.

Em 1949, em co-autoria com o também matemático estadunidense Warren Weaver (1894-1978), publicou o livro Teoria Matemática da Comunicação (The Mathematical Theory of Communication), contendo reimpressões do seu artigo científico de 1948 de forma acessível também a não-especialistas - isto popularizou seus conceitos.

Entre 1946 e 1953, Claude Shannon integrou temporariamente o grupo reunido sob o nome de Macy Conferences, contribuindo para a consolidação da teoria cibernética junto com outros cientistas renomados como John von Neumann e Norbert Wiener. Shannon é famoso por ter fundado a teoria da informação com um artigo publicado em 1948. Mas a ele também é creditado como fundador tanto do computador digital como do projeto de circuito digital em 1937, quando, com 21 anos de idade e aluno de mestrado no MIT, ele escreveu uma tese demonstrando que uma aplicação elétrica utilizando álgebra booleana poderia resolver qualquer problema de lógica. Tem sido dito que esta foi a tese de mestrado de mais importância de todos os tempos. Shannon contribui para o campo da criptoanálise durante a segunda guerra mundial.

Biografia

Shannon nasceu em Petoskey, Michigan. Seu pai, Claude Sr (1862–1934), um descendente dos primeiros colonos de New Jersey, foi um empresário bem sucedido e foi juiz por um certo tempo. Sua mãe , Mabel Wolf Shannon (1890–1945), filha de imigrantes alemães, era uma professora de línguas. Os primeiros 16 anos de Shannon foram em Gaylord, Michigan, onde ele frequentou o ensino público, graduando-se no Gaylord High School em 1932. Shannon mostrou uma inclinação para coisas mecânicas, seus melhores talentos eram para a ciência e matemática. Em casa construiu dispositivos tais como modelos de aviões, um modelo de um barco controlado por rádio e um sistema de telégrafo. Enquanto crescia, trabalhava como mensageiro da Western Union. Seu herói de infância era Thomas Edison, que descobriu depois ser um primo distante. Ambos eram descendentes de John Ogden, um líder colonial e um ancestral de muitas pessoas ilustres.

Teoria Booleana

Em 1932 Shannon começou a cursar a University of Michigan, formando em 1936 com duas graduações de bacharelado em engenharia elétrica e matemática. Posteriormente, começou seus estudos de pós-graduação no Massachusetts Institute of Technology (MIT), onde trabalhou com o analisador diferencial de Vannevar Bush, um computador lógico.

Ao estudar os complexos circuitos ad hoc do analisador diferencial, Shannon viu que os conceitos de George Boole, inventor da álgebra booleana, poderia ser útil para várias coisas. Um documento eleborado a partir da sua tese de mestrado em 1937, A Symbolic Analysis of Relay and Switching Circuits, foi publicado na edição de 1938 da Transactions of the American Institute of Electrical Engineers. Howard Gardner, da universidade de Harvard, chamou a tese de Shannon como "possivelmente a mais importante e também a mais famosa tese de mestrado do século."

Neste trabalho, Shannon provou que a álgebra booleana e a aritmética binária poderiam ser utilizadas para simplificar o arranjo dos relés eletromecânicos e então utilizados em comutadores para roteamento telefônico. Expandindo o conceito ele também mostrou que deveria ser possível a utilização de arranjos de relés para resolver problemas de álgebra booleana. A exploração dessa propriedade de interruptores elétricos criou a lógica e os conceitos mais básico dos computadores digitais. O trabalho de Shannon tornou-se o principal na área de circuitos digitais quando se tornou amplamente conhecido entre a comunidade de engenharia elétrica durante e após a segunda guerra mundial. O trabalho teórico rigoroso de Shannon substituiu completamente os métodos ad hoc que haviam prevalecido anteriormente.

Em 1940, Shannon se tornou pesquisador do Instituto nacional de Estudos Avançados em Princeton, Nova Jersey. Em Princeton, Shannon teve a oportunidade de discutir suas idéias com cientistas e matemáticos influentes como Hermann Weyl e John von Neumann, além de um encontro ocasional com Albert Einstein. Shannon trabalhou livremente em todas as áreas, e começou a moldar as idéias que se tornariam a teoria da informação.

Pesquisa em Tempo de Guerra

Shannon em seguida juntou-se a Bell Labs para trabalhar em sistemas de controle de artilharia e criptografia durante a segunda guerra mundial, sob um contrato com a seção D-2 (Seção de Controle de Sistemas) do Comitê Nacional de Pesquisa em Defesa.

Conheceu sua esposa quando era um analista numerico na Bell Labs. Casaram em 1949.

Durante dois meses no início de 1943, Shannon entrou em contato como o criptoanalista líder e matemático britânico Alan Turing. Turing havia sido enviado para Washington para compartilhar com o serviço de criptoanálise da marinha dos EUA os métodos utilizados pela escola de códigos e cifras do governo britânico em Bletchley Park para quebrar as cifras utilizadas pelos submarinos alemães no Atlântico Norte. Ele também ficou interessado em cifragem de fala e para isso ficou um tempo no Bell Labs. Shannon and Turing se encontraram na hora do lanche em uma cafeteria e Turing mostrou a Shannon seu artigo que definiu o que hoje é conhecido como a "Máquina Universal de Turing". Em 1945, quando a guerra estava chegando ao fim, O NDRC estava emitindo um resumo dos relatórios técnicos como sua última atividade antes de seu eventual fechamento. Dentro do volume de controle de fogo um documento especial entitulado "Suavização de Dados e Previsão em Sistemas de Controle de Fogo",coautoria de Shannon, Ralph Beebe Blackman, e Hendrik Wade Bode, formalmente tratava do problema de suavização dos dados no controle de incêndio por analogia com "o problema de separar um sinal de um ruído interferindo no sistema de comunicação." Em outras palavras foi modelado o problema em termos de dados e processamento de sinal e assim, anunciava o início da era da informação.

Seu trabalho em criptografia foi mais estreitamente relacionada com suas publicações posteriores sobre a teoria da informação. No final da guerra, ele preparou um memorando para a Bell Telephone Labs entitulado "Uma Teoria Matemática da Criptografia," datada de setembro de 1945. Uma versão desclassificada deste trabalho foi posteriormente publicada em 1949 como "Teoria da Comunicação de Sistemas Secretos" no Bell System Technical Journal. Este trabalho incorporou muitos dos conceitos e formulações matemáticas que também apareceram em seu "Uma Teoria Matemática da Comunicação". Shannon disse que suas idéias em teoria da comunicação e criptografia durante a guerra haviam sido desenvolvidas simultaneamente e " elas estavam tão juntas que você não podia separa-las" . Em note de rodapé perto do início do relatório classificado, Shannon anunciou sua intenção de "desenvolver estes estudos... em um memorando sobre a transmissão de informações."

Enquanto na Bell Labs, ele provou que a one-time pad era inquebrável em sua pesquisa que mais tarde foi publicada em outubro de 1949. Ele também provou que qualquer sistema inquebrável deve ter essencialmente as mesmas características do one-time pad: A chave deve ser verdadeiramente aleatória, tão grande quanto o texto original, nunca reutilizada e mantida em segredo.

Contribuições Pós-Guerra

Em 1948, o memorando prometido apareceu como "A Mathematical Theory of Communication", um artigo com duas partes nas edições de julho e outubro do Bell System Technical Journal. Este trabalho enfoca no problema da melhor forma de codificar uma informação que um remetente deseja transmitir. Neste trabalho fundamental ele usou ferramentas da teoria da probabilidade, desenvolvidas por Norbert Wiener, que estavam em seus estágios iniciais de serem aplicadas a teoria das comunicações na época. Shannon desenvolveu a entropia da informação como uma medida de incerteza em uma mensagem.

Contribuição fundamental da teoria da informação para processamento de linguagem natural e lingüística computacional foi ainda estabelecida em 1951, em seu artigo "Previsão e Entropia de Impresso Inglês", mostrando limites superior e inferior da entropia nas estatísticas de Inglês - dando uma base estatística para análise da linguagem.

Outro papel notável publicado em 1949 é a "Communication Theory of Secrecy Systems", uma versão desclassificada do seu trabalho em tempo de guerra sobre a teoria matemática de criptografia , no qual ele provou que todas as cifras teoricamente inquebrável deve ter os mesmos requisitos que a one-time pad. A ele também é creditado a introdução da teoria da amostragem, que se preocupa com o que representa um sinal de tempo contínuo a partir de um conjunto (uniforme) discreto de amostras. Essa teoria foi essencial para permitir a passagem das telecomunicações dos sistemas analógicos para sistemas digitais no ano de 1960 e posteriores.

Hobbies e Invenções

Fora de suas pesquisas acadêmicas, Shannon estava interessado em malabarismo , monociclo e xadrez. Ele também inventou diversos dispositivos. Um dos seus dispositivos mais engraçados era uma caixa mantida em sua mesa chamada de "Máquina definitiva", baseada em uma idéia de Marvin Minsky. Além disso, ele construiu um dispositivo que poderia resolver o Cubo de Rubik. Ele também é considerado o co-inventor do primeiro computador portátil, juntamente com Edward O. Thorp. O dispositivo foi utilizado para melhorar as chances quando se jogar roleta .

Legado e homenagens

Shannon chegou ao MIT em 1956, para se juntar ao corpo docente e para realizar um trabalho no Laboratório de Pesquisa de Eletrônica (RLE). Ele continuou a servir no corpo docente do MIT até 1978. Para comemorar suas conquistas, houve celebrações de seu trabalho em 2001, e atualmente há seis estátuas de Shannon esculpido por Eugene L. Daub : um na Universidade de Michigan , uma no MIT no Laboratório de Sistemas de Informação e Decisão , um em Gaylord, Michigan, um na Universidade da Califórnia, San Diego , uma no Bell Labs , e outro no Labs Shannon AT & T. Após o rompimento na Bell, a parte do Bell Labs que ficou com a AT & T foi nomeado Shannon Labs em sua honra.

De acordo com Neil Sloane , a perspectiva introduzida por Shannon da teoria da comunicação (agora chamada de teoria da informação ) é a base da revolução digital, e cada dispositivo que contém um microprocessador ou microcontrolador é um descendente conceitual da publicação de Shannon ".. Ele é um dos grandes homens do século, sem ele, nenhuma das coisas que conhecemos hoje existiria. Toda a revolução digital começou com ele ".

Shannon desenvolveu a doença de Alzheimer , e passou seus últimos anos em um asilo de Massachusetts. Ele foi auxiliado por sua esposa, Mary Elizabeth Moore Shannon, o filho, Andrew Moore Shannon; a filha, Shannon Margarita, uma irmã, Catherine S. Kay e suas duas netas.

Ludwig Bolzmann



Para iniciar essa discussão precisamos dar um sentido matemático à palavra INFORMAÇÃO ou INCERTEA e definir uma MEDIDA DA INFORMAÇÃO. Para isso o melhor é utilizar a metodologia axiomática na qual definimos certas propriedades que esperamos que a medida possua através dos axiomas e esperamos que os mesmos limitem a classe de funções capazes de expressar a medida, ou, no melhor dos casos, que exista apenas uma função que satisfaz aos axiomas. Claro que, nesse caso, todas as grandezas que satisfezerem aos mesmo axiomas serão agrupadas e chamadas pelo mesmo nome.

Então vamos começar esse capítulo explicitando um conjunto de axiomas que representem algo que definimos como informação. Suponha uma variável aleatória que pode tomar valores no conjunto finito  com probabilidades  com as restrições usuais  e . Queremos associar uma função de  que represente uma medida de incerteza associada à variável aleatória . Vamos construir duas funções  e . A função  será a incerteza associada a um evento de com probabilidade . Se o evento  tem probabilidade  então  é a incerteza associada ao evento . Podemos pensar que  é a incerteza removida ao descobrir que o evento  ocorreu, ou ainda, que é a informação revelada pelo fato de que  ocorreu.

Já a função  é a inceretza média associada ao conjunto dos eventos , logo:



Dessa forma, é a incerteza média removida, ou informação média revelada, por descobrir o valor de . Com essas definições vamos impor axiomas sobre a grandeza INCERTEZA ou INFORMAÇÃO.

**Axioma 1:** Suponha que todos os valores de  são independentyes entre si e igualmente prováveis. Nesse caso  e  será uma função de . Por exemplo, no jogo de cara ou coroa de uma moeda honesta teremos , e no jogo de escolher uma pessoa ao acaso de uma cidade com  habitantes teremos . Aqui cabe a pergunta: o que contém mais informação – descobrir que se obteve coroa em um jogo de cara-ou-coroa ou descobrir que o cidadão José Jesus Pedro (também, com tanta ajuda dos santos) foi o escolhido em uma cidade de 20 milhões de habitantes? Claro que o evento de menor probabilidade, que causa maior surpresa, deve conter mais informação quando revelado. Esse é o primeiro axioma da teoria da informação:

 é monotônica crescente em .

Em outras palavras, se , então .

**Axioma 2:** Considere agora um experimento envolvendo duas variáveis aleatórias independentes. Uma denominada por  com valores possíveis  com iguais probabilidades , e a outra denominada  com valores possíveis  com iguais probabilidades . Como as variáveis são independentes então:



Agora, o fato de ter-se revelado o valor de  não traz qualquer informação sobre o valor de , pois as variáveis são independentes. Assim, a incerteza removida por descobrir o valor de  é apenas a inceretza média de . O mesmo argumento vale para . Assim, se sorteamos primeiro o valor de x extraímos a incerteza  e depois sorteamos  extraindo a incerteza , e no processo inteiro extraímos a incerteza . Agora se sorteamos as duas variáveis simultaneamente, temos 1 possibilidade em , logo extraímos a incerteza . O segundo axioma afirma que é indiferente a ordem de extração da informação, uma antes da outra ou as duas simultaneamente.

Matematicamente isso é escrito como:



**Axioma 3:** Agora vamos remover a restrição de probabilidades iguais e pensar em termos de reagrupamento. Vamos agrupar a variável  em dois conjuntos  e  da seguinte forma:  e . Vamos sortear a variável  com o seguinte processo: primeiro sortear qual grupo,  ou , e sortear  dentro de seu respectivo grupo.

Agora, pelos axiomas da probabilidade:





Daí precisamos responder à pergunta: qual a probabilidade de sortear  sabendo que  foi escolhido? Para responder a essa pergunta usamos os teoremas de Bayes:  (lê-se essa expressão como p de A dado B) onde . Se  então  e . Como  então  então:



Da mesma forma:



Vamos denominar esse experimento de  o processo de escolher primeiro o grupo  e depois o , para diferenciá-lo do experimento escolher o  diretamente. A probabilidade será . Mas  e  logo  como deveria ser, claro.

Quanta incerteza será removida por descobrir qual dos dois grupos foi escolhido?



Dado que o grupo  foi escolhido qual a incerteza remanescente?



A mesma coisa para o grupo :



Assim a incerteza média removida no processo de escolha entre um grupo e outra será:



A incerteza na escolha de um dado  será a incerteza na escolha entre os dois grupos adicionada da incerteza média restante na escolha de  dentro de cada grupo, ou seja:



**Axioma 4:** Simplesmente impõe que  seja contínua em .

Os 4 axiomas da medida da incerteza então são:

1.  é monotônica crescente em.
2. .
3. 
4.  é contínua em .

Com esses 4 axiomas podemos provar o seguinte teorema:

A única função que satisfaz aos 4 axiomas é: . A base do logarítmo não interessa porque pode ser incorporada na constante arbitrária .

Provar que a função dada satisfaz aos axiomas é fácil.

1.  e é uma função monotônica crescente desde que a base seja um número mairo do que 1.
2.  e  logo  logo .
3. 







 CQD

1.  que é uma função contínua no intervalo .

O teorema entretanto afirma mais. Afirma que essa é a única função que satisfaz aos 4 axiomas. Podemos provar o teorema completo da seguinte forma:

1. Mostrar que  por indução.

Primeiro  logo  é verdadeiro. Supor que  é verdadeiro e provar que é verdade para . Pelo axioma 2 temos que  CQD.

1. Mostrar que  com .

Para , pelo axioma 2, temos que:  o que significa que . Logo  é verdadeira para . Para  vamos tomar um inteiro positivo  qualquer. Sempre haverá um . Para provar isso basta aplicar o  em todos os fatores, lembrando que para  então  é monotônica crescente, que nos leva a:

 ou ainda , então  é o antecessor de  e  o sucessor. Pelo axioma 1 sabemos que  e pelo teorema (a) temos que , ou seja:

.

Por outro lado , logo, ou seja:



Podemos subtrair uma equação da outra conisderando os piores casos tanto nos valores positivos quanto nos negativos. Do lado negativo o pior caso é , e no positivo é , assim:

 ou seja 

Como esse resultado é válido para qualquer então fazendo  temos que:  ou que  então:

 CQD

1. Provar que  se  é racional.

Se  é racional então  ode  e  são inteiros positivos e  porque . Agora podemos criar a função com  a partição:



Aplicando o axioma 3 temos:



aplicando (b) obtemos:



Da qual extraímos:



Portanto:

 CQD

Como essa função é válida para todo número racional e a função é contínua então o resultado é válido para qualquer .

1. Finalmente provar que 

Já sabemos que a fórmula é verdadeira para , , , e para , ,  e . Vamos supor que a fórmula é verdadeira para  e mostrar que, então, é verdadeira para . Para mostrar o último ponto precisamos do axioma 3.















Chegando ao resultado final



Ou seja:



Quando Shannon descobriu essa medida ele notou que era idêntica à entropia de Boltzmann:



e a denominou de ENTROPIA, hoje também conhecida como entropia de Shannon.

**Entropia de Shannon e a Física Estatística:**

Princípio da MAXIMIZAÇÃO DA INCERTEZA, ou da ENTROPIA.

Agora nos fazemos a pergunta: qual a distribuição de probabilidades c que maximiza a entropia ?

**Ensemble Microcanônico.** Nesse ensemble só temos a restrição . Nesse caso, o Lagrangeano do sistema é dado por:



A solução será dada por



 ou seja, , a probabilidade é constante, não depende de . Aplicando a restrição vemos que  implica em . Agora, se desejado pode-se determinar o valor do multiplicador de Lagrange .

Essa distribuição é a uniforme, com iguais probabilidades à priori, uma suposição forte da física estatística que diz que, na ausência de qualquer outra informação atribui-se iguais probabilidades. Esse é o caso de ensemble microcanônico, com N partículas fechadas em uma caixa isolada de volume fixo, ou seja, o número de partículas, a energia total e o volume são conhecidos.

**Ensemble Canônico.** Agora existem uma informação extra, sabemos que a energia média do sistema vale . Nosso problema passa a ser achar o máximo da entropia  com as restrições  e . Nesse caso, o Lagrangeano do sistema é dado por:



A solução será dada por



, e .

Encontrando as constantes  e :

 e  portanto 

Ou seja:



Mas da termodinâmica sabemos que , logo  e . Só falta encontrar , aplicando a restrição , que nos leva , ou seja: . Chamamos a função de partição , então  e a probabilidade de cada estado é dada por: . Os físicos gostam de usar o seguinte truque: definir , e escrever  e usar . Esse é a ensemble canônico.

O processo pode ser generalizado incluindo outras restrições. Por exemplo:

Achar o máximo da entropia  com as restrições  ,  e . Nesse caso, o Lagrangeano do sistema é dado por:





 e  logo







As variáveis termodinâmicas conjugadas [uma intensiva e outra extensiva] são dadas por . Então  e  o que significa que , e, finalmente, . Daqui podemos extrair os ensembles grand-canônico e T-P.

**Generalização do Método de Entropia Máxima.**

O problema de inferir qual a distribuição de probabilidades  a partir de ertas observáveis pode ser generalizado colocando o problema como:

Encontrar  que maximiza a entropia

Maximizar  com as restrições:



O Lagrangeano do sistema é dado por:





, . Definindo , , , ... e  as  constantes devem ser extraidas do sistema de equações não lineares das restrições:



Que pode ser re-escrito da forma:



Nesse caso, então, a estratégia é: usar SOLVER para encontrar com as equações:



O valor de é então calculado com .

Caso particular com apenas uma restrição de valores médios:

Encontrar  que seja solução de  e calcular .

Considere um experimento repetido  vezes com  resultados possíveis. Quantos resultados são possíveis? Claro que . Agora nos perguntamos qual a multiplicidade do experimento em que o resultado 1 foi obtido  vezes, o 2 foi obtido  vezes e assim por diante até o resultado , obtido  vezes, de modo que . Multiplicidade significa de quantas formas podemos obter esse resultado.

De quantas formas podemos colocar  objetos de  em uma caixa com primeiro valor do experimento sem reptição? Considere que temos  caixas. Para a primeira caixa temos  opções, para a segunda , pois já tiramos uma opção das , e assim por diante até a caixa  para a qual sobraram  opções, como mostra o diagrama abaixo:



Em princípio teríamos  opções, mas devemos lembrar que a ordem com que os objetos de cada caixa foi retirada não interessa, então temos que dividir esse número pelas permutações possíveis das  caixas. Essas permutações são dadas pelo mesmo problema na forma:  e valem . Logo temos  possibilidades para a primeira caixa. Sobraram  objetos dos quais queremos colocar  na segunda caixa. Repetindo o raciocínio vemos que teremos , na terceira caixa teremos , até a última com . Multiplicando todas elas teremos:



Ou seja:

 com a restrição .

Aplicando o logarítmo neperiano temos: .

A aproximação de Stirling para o logarítmo da função fatorial de números muito grandes [ver notas de série de Taylor] é

, ou seja  logo  e a aproximação de Stirling é dada por: . Usando essa aproximação temos que:





 ou seja:

 e  é a entropia de Boltzmann.