

A IMPORTÂNCIA DA ENTROPIA DA INFORMAÇÃO PARA AS CIÊNCIAS DAS COMUNICAÇÕES

LUCAS ALEXANDRE MONTEIRO BRASIL

ORIENTADOR: DR. CARLOS HANDREY ARAUJO FERRAZ

Resumo: Este trabalho de conclusão de curso apresenta e exemplificar o conceito de entropia da informação. Para isso, foi conduzido uma pesquisa com base no artigo de Claude Shannon, intitulado “*A Mathematical Theory of Communication*” investigando as bases da sua formação teórica. Além disso, foi aprofundado a compreensão da área por meio de análise de artigos e documentários que oferecem perspectivas adicionais sobre o assunto. Com o intuito de proporcionar um entendimento mais abrangente e aplicável, foi conduzido estudo de caso, nos quais é desenvolvido o conceito em cenários do cotidiano. Por fim foi demonstrado como a ideia pode ser aplicado na ciência da comunicação.

Palavras-chave: entropia; informação; Shannon; comunicação.

1. INTRODUÇÃO

A evolução dos meios de comunicação desempenha um papel crucial na forma como o ser humano interage, compartilha informação e molda a sociedade em que vive. Desde os primórdios da comunicação humana, a capacidade de transmitir mensagens de forma eficiente e ampla tem sido fundamental para o progresso e o desenvolvimento da humanidade. À medida que a tecnologia avança os meios de comunicação evoluem em um ritmo acelerado, trazendo consigo um impacto profundo em todas as esferas da vida.

A importância da evolução dos meios de comunicação reside na sua capacidade de conectar pessoas, culturas e informações em escala global. Isso não apenas encurta distâncias geográficas, mas também derruba barreiras linguísticas e culturais. Através dos meios como a imprensa, rádios, televisão e, mais recentemente, a internet com as redes sociais, as sociedades tem a capacidade de se informar, se expressar e se mobilizar como nunca visto antes.

A evolução dos meios de comunicação também desempenha um papel vital no fortalecimento da democracia e na promoção da liberdade de expressão. A disseminação rápida de informação e a capacidade de cidadãos comuns de compartilhar opiniões sobre acontecimentos em tempo real, não ficando vítimas apenas do que é compartilhado pelas imprensas, traz um impacto profundo na formação de opiniões pública e na tomada de decisões políticas. Além disso, os meios de comunicação desempenham um papel fundamental na denúncia de injustiças, na promoção da responsabilidade e da transparência.

Neste contexto, este trabalho de conclusão de curso busca aprofunda a compreensão sobre a relação entre a teoria da informação, especialmente a entropia da informação, e os avanços tecnológicos que moldaram a era digital. Será explorado como a entropia da informação, concebida por Claude Shannon, influenciou a área da telecomunicação. Ao longo desse trabalho, ocorrerá o estudo de casos e exemplos concretos que ilustram a aplicação prática da entropia da informação. Com isso, é esperado obter uma compreensão mais aprofundada de como a teoria da informação, se tornou um pilar fundamental da sociedade digital atual.

2. CLAUDE SHANNON: UM VISIONÁRIO NA TEORIA DA INFORMAÇÃO

Claude Elwood Shannon nasceu nos Estados Unidos da América, na cidade de Petoskey, no estado de Michigan, em 30 de abril de 1916. Desde a infância, demonstrou um profundo interesse pela área da comunicação. Um exemplo desse interesse precoce é um telegrafo que ele construiu usando código Morse quando ainda era criança, na tentativa de se comunicar com a casa vizinha. Embora essa primeira tentativa tenha falhado, ela é um dos primeiros índices do seu interesse na área.

Durante a sua formação acadêmica, obteve dois diplomas de bacharelado um em matemática e outro em engenharia elétrica. Aos 21 anos ele escreve sua tese de mestrado, considerado amplamente uma das mais importantes da história da engenharia elétrica e das ciências da computação, intitulada “*A Symbolic Analysis of Relay and Switching Circuits*”. Na dissertação ele aplica a álgebra booleana de forma criativa e engenhosa a análise de circuitos elétricos. Ele demonstra como as operações lógicas poderiam ser usadas para representar e simplificar circuitos lógicos, tornando-os mais eficientes e práticos.

Em 1948, enquanto trabalhava nos Laboratórios Bell, Shannon publica o artigo “*A Mathematical Theory of Communication*” que inicialmente foi publicando em dois volumes pela “*The Bell System Technical Journal*”. Esses artigos foram onde ele apresentou ao mundo sua teoria revolucionária da informação e suas contribuições para a comunicação, estabelecendo as bases para a teoria da informação moderna.

Shannon tinha a preocupação de quantificar o quanto de informação pode ser transmitido através de um canal de comunicação sujeito a ruídos, para isso ele teve a ideia de transformar toda a informação em bits, logo convertendo-as em zeros e uns. Nesse artigo, foi introduzido o termo bit, como também a entropia da informação, que será explicada no tópico 3 e como ela influenciou as diversas áreas até os dias atuais.

O artigo, começou a ser aplicado nas mais diversas áreas, tendo a relevância bem maior do que ele podia esperar, o fato dele propor uma comunicação segura, com os mínimos de erros possível, chamou atenção, se transformando em uma área de estudo de diversos cientistas, colocando assim Shannon entre as principais referências para a área de comunicação.

No artigo, para além dos pontos previamente mencionados, Claude Shannon avança ainda mais em seu trabalho, apresentando o teorema da codificação de canal. Este teorema, crucial em sua importância, estabelece limites fundamentais para a transmissão de informações em canais de comunicação que estão sujeitos a ruídos e interferências. Além disso, Shannon introduz um outro elemento notável: o diagrama do modelo de comunicação. Este diagrama, que posteriormente se tornaria um padrão na descrição de sistemas de comunicação, desenha o processo de comunicação de maneira clara e concisa. Esse modelo inclui componentes essenciais como a fonte de informação, o transmissor, o canal e o receptor, proporcionando uma estrutura sólida para entender e analisar sistemas de comunicação complexos.

Em 1956 ele aceita uma posição de professor na *Massachusetts Institute of Technology (MIT)* onde deu continuidade as suas contribuições significativas no campo da teoria da informação. Shannon permaneceu no MIT até sua aposentadoria em 1978.

Outras áreas chamavam a atenção de Shannon como por exemplo a ideia de máquinas inteligentes. Durante estudos elaborou diversas máquinas, sendo um dos pioneiros na área de inteligência artificial, elaborou um rato mecânico, que conseguia aprender o caminho no labirinto, aprendendo com os seus erros. Outra a ser citada é uma máquina que joga xadrez, tendo construído um dos primeiros computadores a jogar o jogo, não tinha capacidade de jogar o jogo inteiro, por causa das limitações tecnológicas da época, mas conseguia calcular os últimos 6 movimentos de um jogo, que ganhou o nome de “*End game*”.

Durante seus últimos anos de vida Shannon enfrentou problemas de saúde, incluindo a doença de Alzheimer. Infelizmente, essa condição afetou sua capacidade de se comunicar e lembrar, o que foi especialmente triste para alguém que contribuiu tanto para a teoria da informação e comunicação. Claude Shannon faleceu em 24 de fevereiro de 2001, em Medford, Massachusetts, deixando para trás um legado duradouro e influente no campo da ciência da computação, engenharia de comunicação e inteligência artificial. Seu trabalho continua a ser uma fonte de inspiração para cientistas e pesquisadores em todo o mundo.

3. O QUE É A ENTROPIA DA INFORMAÇÃO:

Shannon apresenta a entropia da informação como uma medida que se quantificar uma incerteza ou imprevisibilidade que uma mensagem pode apresentar. Em essência, ela ajuda a compreender quão caótico ou organizado um conjunto pode ser. Quando a entropia é alta isso significa que o sistema está repleto de incerteza, e a informação é altamente imprevisível, tornando difícil antecipar o próximo resultado ou mensagem. Porém quando o contrário acontece e se tem a mínima entropia, logo se tem um texto organizado e previsíveis, se tendo um mínimo de incertezas possível.

No seu artigo, Shannon optou por utilizar os logaritmos para calcular a entropia por melhor representar os resultados desejados por ser útil, fiel e matematicamente mais adequada. A escolha da base depende da unidade de medida para a informação, se for de base 2 é obtido os dados binários ou no caso os bits. Ele apresenta a seguinte equação, como o método que quantifica a informação.

$$H(x) = - \sum [P(x) \cdot \log_2(P(x))]$$
 (1)

Aqui, $H(x)$ é a representação escolhida para a entropia da fonte, que é igual ao somatório das probabilidades de ocorrência de cada termo multiplicadas pelo logaritmo na base 2 da recíproca de $P(x)$. Por uma propriedade do logaritmo de uma divisão, podemos simplificar a equação adicionando o sinal de menos, como visto na equação. No item a seguir serão apresentados casos e como a aplicação da entropia funciona.

3.1. ESTUDO DE CASOS

Após a compreensão teórica desses conceitos fundamentais, é essencial dar um passo adiante e explorar sua aplicação em exemplos práticos e a análise de resultados concretos. A teoria fornece o arcabouço conceitual, mas é na implementação prática que vemos como essas ideias se traduzem no mundo real. A seguir será apresentado exemplos de situações que todas as possibilidades possuem a mesma probabilidade de ocorrer, seguido pela discussão do oposto dessas situações, ou seja, quando as probabilidades variam.

3.1.1. QUANDO CADA SITUAÇÃO TEM A PROBABILIDADE IGUAL

Primeiro exemplo a ser citado vai ser o sorteio da escolha de lados organizados durante um jogo de futebol, onde inicialmente o juiz chama os capitães de ambos os times e com um auxílio de uma moeda, o time A fica com a face da moeda considerada “cara” enquanto o time B fica com a “coroa”, que é o lado oposto, considerando que a moeda seja legal e que nenhuma de suas faces seja manipulado, para alteração de resultado, é esperado que cada face da moeda tenha cinquenta por cento de chance cada de sair, mesmo ao analisar o resultado de cinquenta

a cem lançamentos é totalmente improvável ter total certeza qual será o próximo resultado. Neste caso a entropia seria máxima e ambos os times possuem a mesma chance de escolher o lado que deseja.

Ao aplicar a frequência de ocorrência de cada face de uma moeda em um lançamento, é possível obter a informação desejada com apenas uma pergunta. Por exemplo, ao perguntar "O resultado é cara?", se a resposta for sim, obtemos a informação desejada. Caso contrário, a única opção restante é coroa.

No entanto, à medida que aumenta o número de resultados possíveis, o número de perguntas necessárias também aumenta. Para melhorar a eficiência desse processo e evitar a necessidade de avaliar cada situação individualmente, é possível agrupar os casos e utilizar estratégias para reduzir o número de perguntas necessárias. Isso permite uma abordagem mais eficiente na obtenção da informação desejada.

Analisando um sistema que possui 4 resultados possíveis, A, B, C ou D todos com vinte e cinco por cento de acontecer é possível obter o seguinte resultado:

$$H(x) = -[0,25 \cdot \log_2(0,25) + 0,25 \cdot \log_2(0,25) + 0,25 \cdot \log_2(0,25) + 0,25 \cdot \log_2(0,25)] \quad (2)$$

$$H(x) = 2$$

Logo apenas com 2 perguntas é possível obter o resultado desejado. Como por exemplo "O resultado foi A ou B?" se sim logo com mais uma pergunta é possível se ter a informação, o mesmo acontece se o resultado for não, iria ficar sobrando só C ou D.

No entanto, quando se trata de uma moeda em que ambas as faces são idênticas, a entropia resultante de um lançamento é igual a zero. Isso ocorre porque o resultado é completamente previsível, com a certeza de cem por cento de que, em qualquer número de lançamentos, o resultado será invariavelmente o mesmo.

3.1.2. QUANDO CADA SITUAÇÃO TEM A PROBABILIDADE DIFERENTE

O primeiro caso dessa situação a ser analisado vai ser a da própria utilização do alfabeto na produção de textos na língua portuguesa. O alfabeto latino é composto de 26 letras, em um caso onde teria que identificar uma letra através de perguntas de sim ou não, como visto no ponto anterior, cada letra possuiria 3,84% de chance de ser a escolhida, aplicado a equação seria possível obter o seguinte resultado.

$$H(x) = -26 \cdot [0,0384 \cdot \log_2(0,0384)] \quad (3)$$

$$H(x) \cong 4,70$$

Isso implica que seriam necessárias aproximadamente 5 perguntas para se obter o resultado desejado. Entretanto, no segundo caso, quando o sorteio é realizado com base em um trecho de um livro em que cada letra tem a mesma frequência de ocorrência, a entropia seria consideravelmente diferente do cenário anterior. Para um melhor entendimento, será apresentado um trecho do livro "500 Dias Sem Você," escrito pela autora Samantha Silvano.

"A gente precisa estar disposto pra fazer dar certo, pra não se apegar ao passado, pra controlar nossas expectativas, pra aprender com os próprios erros e, sobretudo, pra desistir. Porque, às vezes, nosso coração precisa ter paz, e não ter alguém."

Nesse contexto, uma tabela será elaborada com a inclusão de cada letra, sua frequência de ocorrência e a probabilidade atribuída a cada letra. Essa abordagem possibilitará a análise da organização da informação no texto e a determinação da entropia associada. A entropia auxiliará na compreensão do grau de previsibilidade ou caos existente na escolha das letras e palavras empregadas.

Tabela 1- Frequência de letras

Letra	Frequência	Probabilidade (%)
A	28	14,29
B	1	0,51
C	8	4,08
D	6	3,06
E	24	12,24
F	1	0,51
G	3	1,53
H	0	0
I	7	3,57
J	0	0
K	0	0
L	2	1,02
M	2	1,02
N	7	3,57
O	22	11,22
P	16	8,16
Q	1	0,51
R	26	13,27
S	22	11,22
T	11	5,61
U	3	1,53
V	2	1,02
W	0	0
X	1	0,51
Y	0	0
Z	3	1,53

Analisando os resultados foi possível obter que a frequência da letra “A” é superior que as demais com 28 repetições, enquanto consoantes como o “J” e o “H” não aparecem no texto. Analisando por meio da equação é obtido:

$$H(x) = -[(0,1429 \cdot \log_2 0,1429) + 4 \cdot (0,0051 \cdot \log_2 0,0051) + (0,0408 \cdot \log_2 0,0408) + (0,0306 \cdot \log_2 0,0306) + (0,1224 \cdot \log_2 0,1224) + 3 \cdot (0,0153 \cdot \log_2 0,0153) + 2 \cdot (0,0357 \cdot \log_2 0,0357) + 3 \cdot (0,0102 \cdot \log_2 0,0102) + 2 \cdot (0,1122 \cdot \log_2 0,1122) + (0,0816 \cdot \log_2 0,0816) + (0,1327 \cdot \log_2 0,1327) + (0,0561 \cdot \log_2 0,0561) + 5 \cdot (0 \cdot \log_2 0)] \quad (4)$$

$$H(x) \cong 3,71$$

Logo foi possível obter o resultado em praticamente 4 perguntas, 1 a menos que a anterior, isso se deve ao fato de que as letras sem frequência foram excluídas da equação, reduzindo o conjunto em cinco letras. Por meio da formulação de perguntas bem planejadas, levando em consideração as probabilidades associadas a cada letra, resultados podem ser alcançados de maneira mais eficaz.

Uma estratégia eficaz envolve começar questionando se a letra escolhida está entre as letras “A”, “E”, “O”, “R” e “S”. A soma das probabilidades dessas letras é de 62,24%. Se a resposta for afirmativa, restará um conjunto de cinco letras. Com mais duas perguntas bem elaboradas, será possível identificar a letra escolhida.

Entretanto, se a resposta for negativa, o conjunto restante consistirá em dezesseis letras, após a exclusão daquelas com frequência nula. Mesmo nessa situação, é viável empregar o mesmo método para criar um subconjunto mais restrito que facilite a escolha e reduza a incerteza.

Se a letra “A” possuíse uma frequência de 50%, a pergunta mais eficiente seria isolar a letra, pois se sim rapidamente já mataria as perguntas, porém se não restaria o conjunto com as outras vinte e cinco letras, analogamente deve escolher o mínimo de letras possíveis que representem os outros 25% da frequência, assim em diante.

4. A INFLUENCIAS DA ENTROPIA NA TELECOMUNICAÇÃO

O conceito de entropia da informação, tiveram um impacto significativo na área da telecomunicação. A entropia é uma medida de incerteza da desordem em um conjunto de dados ou mensagens. Ela é fundamental para entender como a informação é transmitida, comprimida e processada em sistemas de comunicação.

Shannon em seu artigo afirma que o problema fundamental da comunicação é o de reproduzir uma mensagem em um ponto, de tal forma que a reprodução dela seja exatamente ou aproximadamente, em outro ponto. Logo ao programar um sistema, o mesmo deve operar para cada seleção possível, não apenas na qual é realmente escolhido, pois no momento em que está projetando é desconhecido. No sistema há várias mensagens que podem ser transmitidas ou recebidas. Estas mensagens podem variar em conteúdo, formato, tamanho, etc. Logo o sistema

tem que identificar a mensagem real e transmiti-la.

Partindo deste ponto, a entropia vai influenciar na área da compressão de informação, demonstrando como quantificar informações de uma fonte, codificando adequadamente as sequências de mensagens em sinais. Como visto nos estudos de casos, quando se tem situações com mesma probabilidade, a representação e codificação da mensagem ficaria bem mais complexa, pois as letras apresentaria incertezas durante a transmissão, porém como visto na linguagem cotidiana, as letras possuem diferentes frequências entre si, logo é possível utilizar de símbolos mais simples para representar letras que tem mais repetição, pelo outro lado quanto as que tem menos repetições podem ser representadas de forma mais longas. Economizando assim espaço de armazenamento ou largura de banda de transmissão.

Após a codificação da mensagem, a mesma é submetida a transmissão através de um canal de comunicação, por exemplo sinais elétricos ou através de cabos, Shannon utilizasse da entropia para calcular a capacidade que um canal tem de transmitir uma quantidade máxima de símbolos por unidade de tempo, tendo uma taxa de erro aceitável. Isso auxilia no projeto de sistemas de comunicações eficazes, garantindo que a comunicação tenha o mínimo de erros possíveis.

Mesmo com a codificação da mensagem, a mesma ainda está sujeita a ruídos, mesmo que as mensagens estejam em dígitos binários e só reconheça dois tipos de intensidade. A mensagem original pode acabar sendo modificada durante o percurso, cabe ao receptor ao receber a mensagem, decodificá-la e submetê-la a correção de erros, a fim de que se tenha a mensagem original. Por fim a mensagem é submetida ao destinatário. Com isso Shannon propõe o diagrama representado na figura 1.

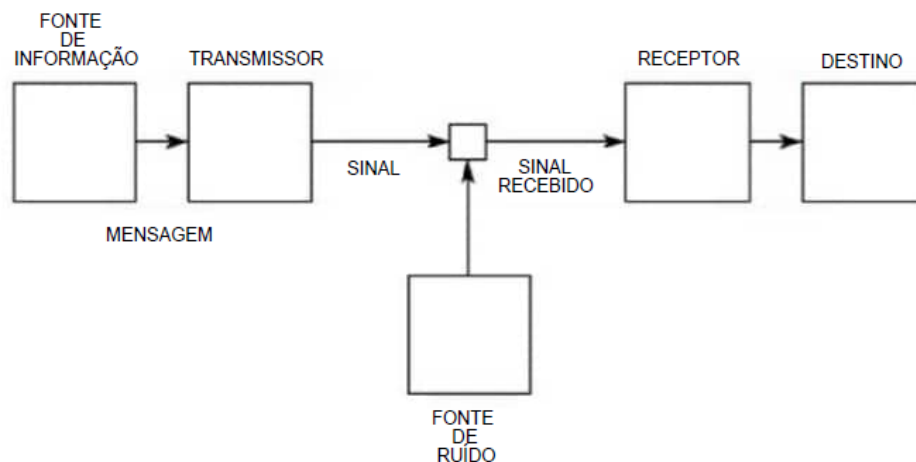


Figura 1- Diagrama esquemático de um sistema de comunicação (A Mathematical Theory of Communication)

Outros pontos a ser utilizado a entropia é na criptografia de senhas e chaves, onde devem ser geradas com máxima entropia possível, para que seja difícil de adivinhar as senhas, evitando a facilitar que adivinhem a mesma. A entropia é usada para medir a incerteza associada a uma chave e sua resistência a ataques de força bruta.

5. CONCLUSÕES

Com os resultados obtidos no estudo foi possível entender que o desenvolvimento da entropia da informação desempenhou um papel crucial na formação do mundo moderno. Entendendo o seu conceito, é possível compreender como uma mensagem é transmitida de forma segura e confiável, sendo um critério essencial em uma sociedade cada vez mais orientada pela informação.

Ao longo deste estudo foi possível entender a importância da entropia como uma medida de incerteza e imprevisibilidade na transmissão de informações. Foi possível compreender como combinações binárias, possibilita a obtenção de informação de forma precisa e eficiente.

Aprofundando os estudos direcionados a compactação e segurança de informação, possibilita entender as raízes da era digital e como novos dispositivos e tecnologias emergem a cada dia, interconectando as sociedades.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- PELEGRINI, C.H. Claude Elwood Shannon e a *A Symbolic Analysis of Relay and Switching Circuits*: Tornando o computador uma máquina semiótica. Contemporânea, Brasil. v. 7, n.1, Jun, 2009
- PINEDA, J.O.C. A entropia segundo Claude Shannon: O desenvolvimento fundamental da teoria da informação: 2006. 9-88p. Dissertação (Mestrado)- Pontifica Universidade Católica, São Paulo
- SHANNON, Claude. *A Mathematical Theory of Communication*. Nova York: *The Bell System Technical Journal*, 1948. 379–423, 623–656 p. v. 27.
- THE BIT PLAYER 2018 – CLAUDE SHANNON. Direção: Mark Levinson. Produção: Mark Levinson. Estados Unidos: *IEEE Information Theory Society* , 2018. Youtuber