

Complexidade e Teoria da Informação: Obtendo o "Código Morse Brasileiro"

Complexity and Information Theory: Obtaining the "Brazilian Morse Code"

RESUMO

Wagner Pinto de Lima
wagnerdelima2@gmail.com
Universidade Tecnológica Federal do Paraná, Pato Branco, Paraná, Brasil

Dr. Angel Akio Tateishi
angeltateishi@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Pato Branco, Paraná, Brasil

Se o Código Morse fosse criado nos dias atuais ele seria o mesmo? E se o Código Morse fosse baseado no idioma Português, como ele seria? Para responder tais perguntas é necessário entender como surgiu o código morse. Neste trabalho será apresentado como foi feito, em 1836, o estudo para a escolha do Código Morse relacionado a cada caractere. Reproduzimos esse estudo desenvolvendo um software, utilizando a linguagem Python, que faz a análise a partir de base de dados com textos de um qualquer idioma que utiliza o alfabeto Latino. A partir da análise estatística da frequência de letras cria-se o respectivo Código Morse para o idioma da base de dados. E por fim, o sistema faz uma comparação entre a versão do Código Morse internacional e a versão gerada pelo software, oferecendo assim uma ferramenta útil para verificar o quão mais simples o Código morse pode se tornar para o país objeto de estudo.

PALAVRAS-CHAVE: Código Morse. Comunicação. Python. Telégrafo.

Recebido: 19 ago. 2019.

Aprovado: 01 out. 2019.

Direito autoral: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



ABSTRACT

If one created the Morse Code today, would it be the same? And if the Morse Code were based on the Portuguese language, what would it look like? To answer such questions, we must understand how the Morse Code was designed. In this paper, it will be presented how it was done, in 1836, the study for the choice of the Morse Code related to each character. We reproduce this study through software, developed in Python language, which analyzes a database of texts of any language based on the Latin alphabet. From the statistical analysis of the frequency of the letters, the respective Morse Code for the database language is created. Finally, the system compares the international Morse Code version with the software-generated version, providing a useful tool for verifying how simpler the Morse Code can become for the country under study.

KEYWORDS: Morse Code. Communication. Python. Telegraph.

INTRODUÇÃO

A *entropia* é um dos conceitos mais importantes da Física. A conexão desse parâmetro macroscópico com a dinâmica microscópica dos inúmeros constituintes de um sistema físico é conhecida como *entropia de Boltzmann-Gibbs*; um dos resultados mais importantes da Mecânica Estatística. Ao desenvolver as bases matemáticas da *Teoria da Informação*, Claude Shannon obteve uma expressão análoga à entropia da Mecânica Estatística, a qual foi denominada *entropia de Shannon*. Dessa forma todos os conceitos e métodos utilizados na física de fenômenos estocásticos podem ser utilizados no estudo da teoria da informação. Por exemplo, o comportamento macroscópico de um gás é caracterizado como sendo completamente aleatório e desorganizado (alta entropia), o qual corresponde aos microestados mais prováveis do sistema. Um sistema organizado e de alta complexidade teria uma probabilidade extremamente pequena de ocorrer sem que energia externa seja utilizada para organizar o sistema. Um sistema de alta complexidade é a comunicação humana, cuja linguagem escrita é baseada em poucos símbolos (letras do alfabeto), mas que é capaz de expressar os mais complexos pensamentos da mente humana. Juntar letras e palavras aleatoriamente é algo que gasta pouca energia, porém não produz nenhum significado útil (apesar de ser uma técnica muito utilizada em alguns trabalhos acadêmicos e em provas nas universidades); já formar sentenças, parágrafos e textos coerentes requer muito esforço e dedicação.

Nesse contexto, aqui serão apresentados alguns dos resultados do plano de trabalho "**Complexidade e Teoria da Informação: Métodos Computacionais e Desenvolvimento Histórico**", relacionado projeto "**Fenômenos Estocásticos e Difusão Anômala em Sistemas Complexos**". Os objetivos principais do plano de trabalho [i) Estudar conceitos e métodos utilizados para analisar a complexidade de um sistema a partir de seus dados (sendo a parte computacional baseada na linguagem de programação Python); ii) Estudar o desenvolvimento histórico dos meios de comunicação baseados em eletromagnetismo, para tentar compreender como informação é codificada e transmitida e como esse processo ficou cada vez mais complexo com os avanços científicos e tecnológicos] foram contemplados concomitantemente no estudo da estrutura do Código Morse.

O famoso código "dot-dash" (ponto-traço) introduzido por Samuel F. B. Morse possui duas importantes características que pouco são mencionadas: é um código binário ou seja um código digital, possui um aspecto estatístico utilizado pois Morse o desenvolveu de tal maneira que as letras mais utilizadas fossem associadas com os símbolos mais curtos (Cherry,1951). Mais conhecida é a relação do Código Morse e os telegrafos. O primeiro telégrafo funcional tinha vinte e seis fios, um para cada letra do alfabeto. Isso o tornava praticamente inviável para a época. Posteriormente inventores alemães conseguiram um telégrafo com cinco fios, porém o pioneiro a fazer um telégrafo com apenas um fio foi Samuel Morse, por volta de 1836 (SALGADO, 2000). Portanto, nota-se que o primeiro projeto do telégrafo não era viável. O primeiro telégrafo estabeleceu comunicação entre as cidades de Washington e Baltimore, no ano de 1837, um

ano após o senado americano investir \$ 30.000,00 dólares no projeto de Morse (equivalente nos dias de hoje a quase \$ 500.000,00 dólares) (BODANIS, 2008). O código Morse é composto por pontos e traços, passando por várias mudanças até chegar no código que está representado na figura 1. De acordo com Salgado (2000), a maior mudança ocorreu em 1948, onde houve onze alterações e passou a ser chamado de Código Morse Internacional.

Figura 1 – Código morse internacional

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

Fonte: <https://s1.static.brasilecola.uol>

Como aplicação e motivação do contexto mencionado, um objetivo particular do nosso trabalho foi responder a seguinte pergunta: *Se o código Morse você baseado em textos do idioma Português, esse código seria igual ao código Morse internacional?* Portanto, desenvolvemos um procedimento para obter o Código Morse para países que utilizam o alfabeto Latino. Para realizar a inferência estatística da ocorrência de letras, utilizamos letras de músicas. Com base nesses textos o sistema irá processar as informações e gerar um Código Morse otimizado para o idioma do país escolhido. Com posse nessas informações será codificado uma mensagem e posteriormente comparada as duas versões: a versão Morse internacional e a versão Morse gerada pelo sistema para aquele país. Na sequência, apresentaremos os métodos desenvolvidos na linguagem de programação python, seguido da apresentação e discussão dos resultados obtidos ao aplicarmos esse métodos para gerar o Código Morse nos idiomas selecionados.

MATERIAIS E MÉTODOS

A fim de aperfeiçoar o processo de análise das letras mais utilizadas em qualquer país, o qual seja objeto de estudo, foi desenvolvido um sistema WEB o qual tem a função de cadastrar textos separados por países. Na presente pesquisa, foi optado por utilizar letras de músicas, porém é possível utilizar qualquer forma de texto.

Para cadastrar um texto no sistema, de forma obrigatória o usuário deve informar o país o qual está sendo estudado. Dessa forma, o software desenvolvido é aplicado nos textos escolhidos para formar a base de dados, verificando a frequência de ocorrência de cada letra do alfabeto e dos números de 0 até 9. Com base nesses cadastros o software irá fazer o processamento de tal maneira que associe as letras mais frequentes com os símbolos (ponto-traço) mais simples, gerando como saída o Código Morse para o idioma de determinado país. Esse resultado é mostrado de duas formas distintas: Gráfico de frequência de letras e Tabela com código Morse.

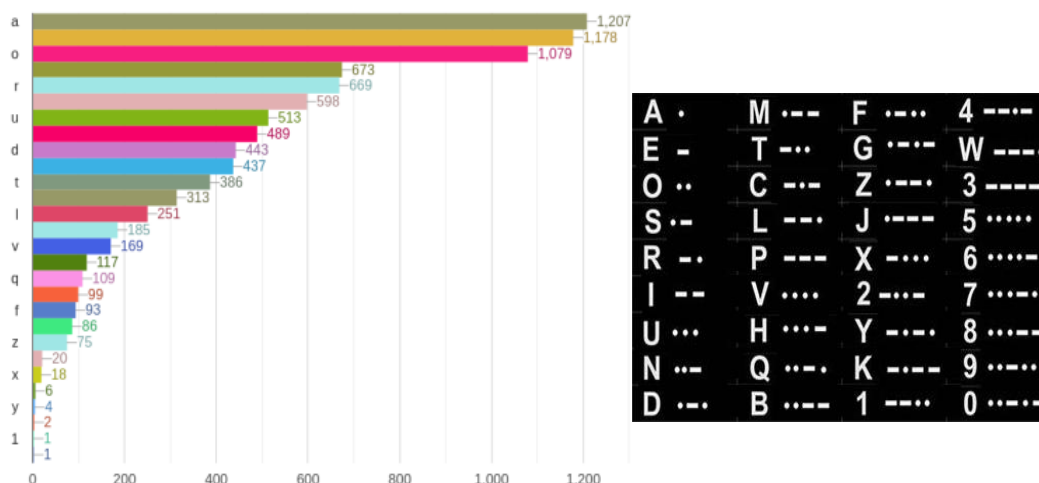
O sistema desenvolvido utilizou Python na versão 3.7.4 e o framework Django na versão 2.2. Para o banco de dados foi utilizado o MySQL na versão 8.0.17. Para a interface gráfica do sistema foi utilizado HTML 5, Javascript e JQuery e o framework Materialize. O servidor foi utilizado uma máquina com sistema operacional Linux, distribuição Ubuntu na versão 18.04 LTS com processador CORE I5 e 6 Gb de memória.

RESULTADOS E DISCUSSÃO

Neste estudo foi utilizado como base os idiomas Inglês (Americano) e o Português (Brasil) e entende-se que cada caractere do Código Morse equivale a um *bit* de informação. Também desconsideram-se os espaços entre palavras, pois é utilizado como padrão de envio o modelo de Ritmo e Representação.

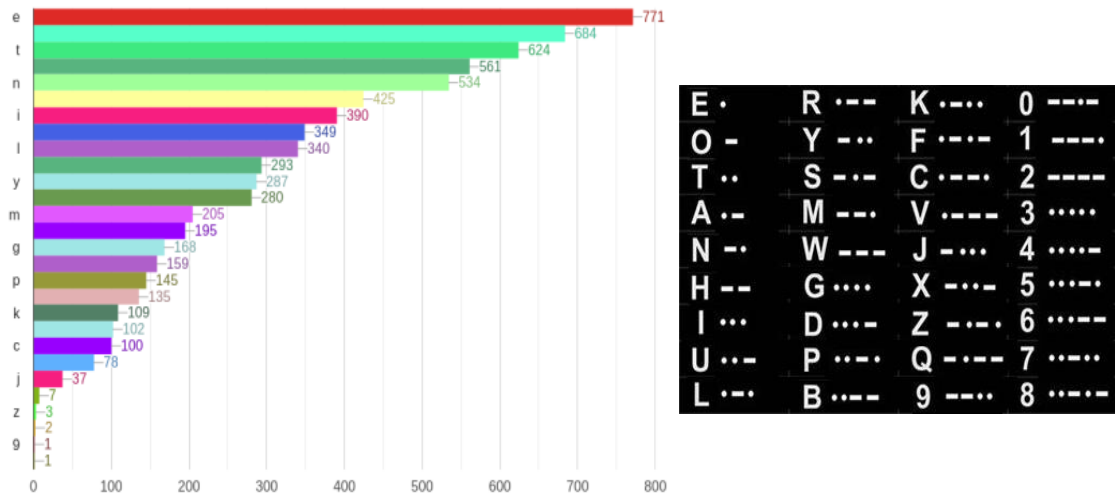
Dessa forma, os resultados obtidos para o idioma Português (Brasil) estão representados na Figura 2, onde é possível reparar que as vogais estão entre as letras mais frequentes, e portanto associadas com símbolos mais simples. Além disso, com a finalidade de verificar a eficiência do método desenvolvido, analisamos textos no idioma Inglês (Americano). Os resultados obtidos estão representados na Figura 3.

Figura 2 – Frequência de letras mais utilizadas e respectivo Código Morse – Português (Brasil)



À esquerda o histograma de frequência de letras mais utilizadas no idioma Português (Brasil). A partir desses dados o sistema gerou o "Código Morse Brasileiro". Fonte: Autoria própria (Software)

Figura 3 – Frequência de letras mais utilizadas e respectivo Código Morse - Inglês (Americano)



À esquerda o histograma de frequência de letras mais utilizadas no idioma Português (Brasil). A partir desses dados o sistema gerou o "Código Morse Brasileiro". Fonte: Autoria própria (Software)

A partir desses resultados, é possível verificar que houve uma grande economia de bits para o idioma português, já para o idioma inglês, houve pouca alteração. Em outras palavras, utilizar o "Código Morse Brasileiro" aqui obtido exigiria o uso de menos informações do que se fosse utilizar o Código Morse Internacional. Pois o "Código Morse Brasileiro" foi obtido baseado em aspecto estatístico do idioma Português. Para verificar esse fato, no idioma Português (Brasil) foi solicitado ao sistema para comparar a palavra "SOCORRO", utilizando o código morse internacional e o código morse gerado pelo próprio sistema, ou seja, aquele otimizado. Como pode ser visto na Figura 4, houve uma economia de sete bits, resultado bem satisfatório, uma vez em que a economia é próxima a um byte (oito bits). Já no idioma Inglês (Americano), o resultado apresentou poucas alterações, no quesito de economia de bits. Porém, essa pequena diferença de bits evidencia que o código gerado por Samuel Morse, em 1836, é muito parecido com o código que o software gerou no ano em que esse trabalho foi produzido, isto é, 2019. Como pode ser visto na figura 4. Para tal análise, foi solicitado ao sistema para comparar a palavra "SOS". A economia foi de apenas dois bits, ou seja, o código gerado pelo software se mostrou similar ao código criado em 1836 por Samuel Morse, o qual passou por onze alterações em 1948.

Figura 4 – Resultado da comparação entre o total de bits utilizados para codificar uma palavra.



À esquerda a palavra "SOCORRO" no sistema padrão e no código obtido para o Português. À direita o termo "SOS" no sistema padrão e no código obtido para o Inglês. Fonte: Autoria própria (Software)

CONCLUSÃO

Dentro do contexto do estudo da Complexidade e da Teoria da Informação, este trabalho de iniciação científica teve objetivos básicos: o aprendizado da linguagem de programação *Python* para ser utilizado como ferramenta para desenvolver algoritmos para o estudo da complexidade de sequências simbólicas (nesse caso letras e números); revisão histórica do desenvolvimento dos meios de comunicação e da teoria da informação. Como uma aplicação e motivação desses estudos foi desenvolvido um software para criar o respectivo Código Morse para qualquer idioma que utiliza o alfabeto Latino. Em particular, obtivemos o "Código Morse Brasileiro" e argumentamos que esse código gastaria menos bits para transmitir mensagens do que quando se utiliza o Código Morse Internacional para codificar palavras em Português. Além disso, também construímos um "telégrafo digital", cuja automação conta com uma placa Arduino Uno e a sua programação foi desenvolvida na linguagem C. Este equipamento possui um display de cristal líquido, para mostrar ao usuário o que está sendo feito, possui também um botão, o qual é a única forma de entrada de dados a qual interpreta comandos em código morse. Há também um led 5mm que auxilia na entrada de dados, ou seja, toda vez que o usuário informa o código morse que representa um caractere, ele deve aguardar o led acender para inserir o próximo código. Considerando o projeto de iniciação científica voluntária teve início em Novembro de 2018, acreditamos que os resultados obtidos foram satisfatórios de acordo com o planejamento inicial. Por fim, esperamos que a leitura e os resultados deste trabalho não tenham correspondido a um sistema de alta entropia.

AGRADECIMENTOS

Agradeço ao professor Doutor Angel Akio Tateishi meu orientador, por confiar a mim esse desafio, foi de muita aprendizagem tanto acadêmica quanto pessoal. Agradeço também aos meus pais que sempre me apoiaram e incentivaram.

REFERÊNCIAS

CHERRY, E. C. **A history of the theory of information**. Proceedings of the IEE - Part III: Radio and Communication Engineering, 98(55), 383–393. doi:10.1049/pi-3.1951.0082 (1951).

SALDAGO, H. D. M. **Código Morse: o que é e como surgiu**. Departamento de Engenharia Informática – Universidade de Coimbra. Coimbra, 2000. Disponível em: <https://student.dei.uc.pt/~hsalgado/CP/artigo.htm>. Acessado em: 03 de agosto de 2019.

BODANIS, D. **Universo Elétrico**. Editora Record. Rio de Janeiro, 2008.