

Shannon-Fano Algorithm

Michelle Orru



...



Topics Covered

INDEX

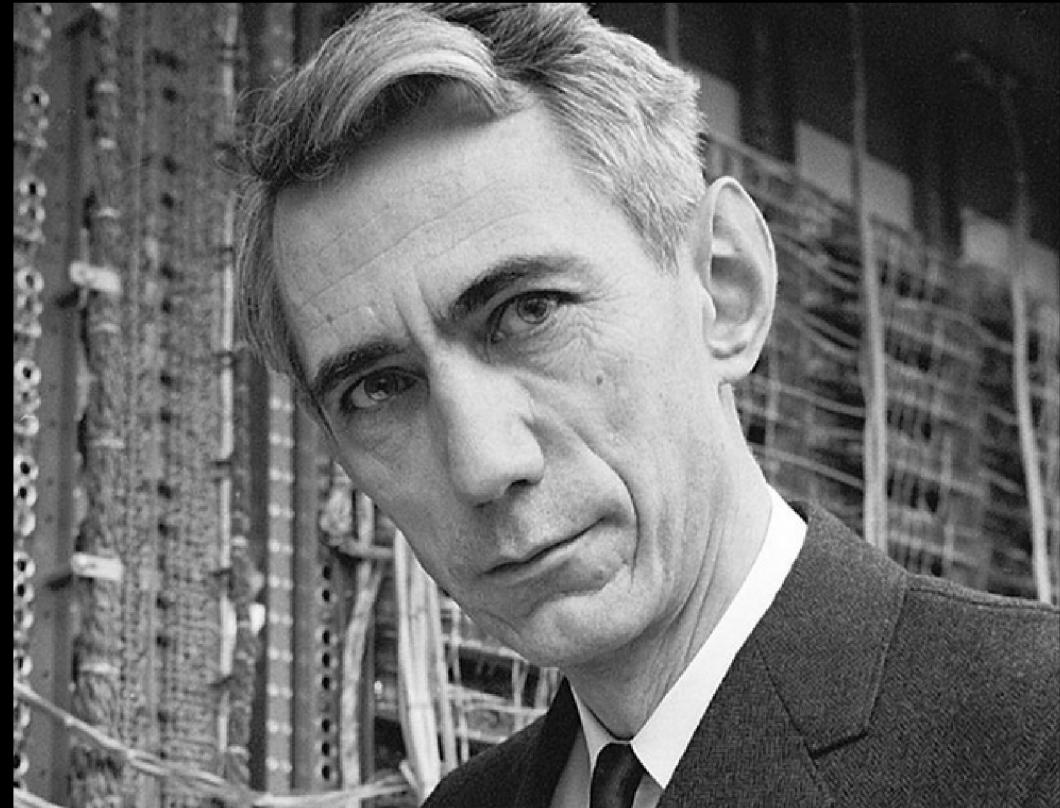


- 01 • Introduction
- 02 • Category
- 03 • Description
- 04 • Purpose
- 05 • Complexity



Background

- The Shannon-Fano algorithm was developed by Claude Shannon and Robert Fano in the late 1940s
- This algorithm was used as a data compression coding technique
- It uses the probabilities of the repetition of characters and assigns a unique code to them
- The least frequent characters get assigned longer codes and the most repeated ones gets a shorter code



CLAUDE SHANNON



ROBERT FANO

Category of the Algorithm

CATEGORY: RECURSIVE ALGORITHM

The Shannon-Fano algorithm is classified as a recursive algorithm. It operates by repeatedly splitting the set of symbols into two subgroups, until each subset consists of just one subtree. The algorithm calculates ...

A	B	C	D	E
0.22	0.28	0.15	0.30	0.05
00	01	10	110	111

... the probability of each symbol in the subset at each step of the recursion and then assigns a binary code to each symbol. This is constructed by appending a "0" or "1" to the assigned character.



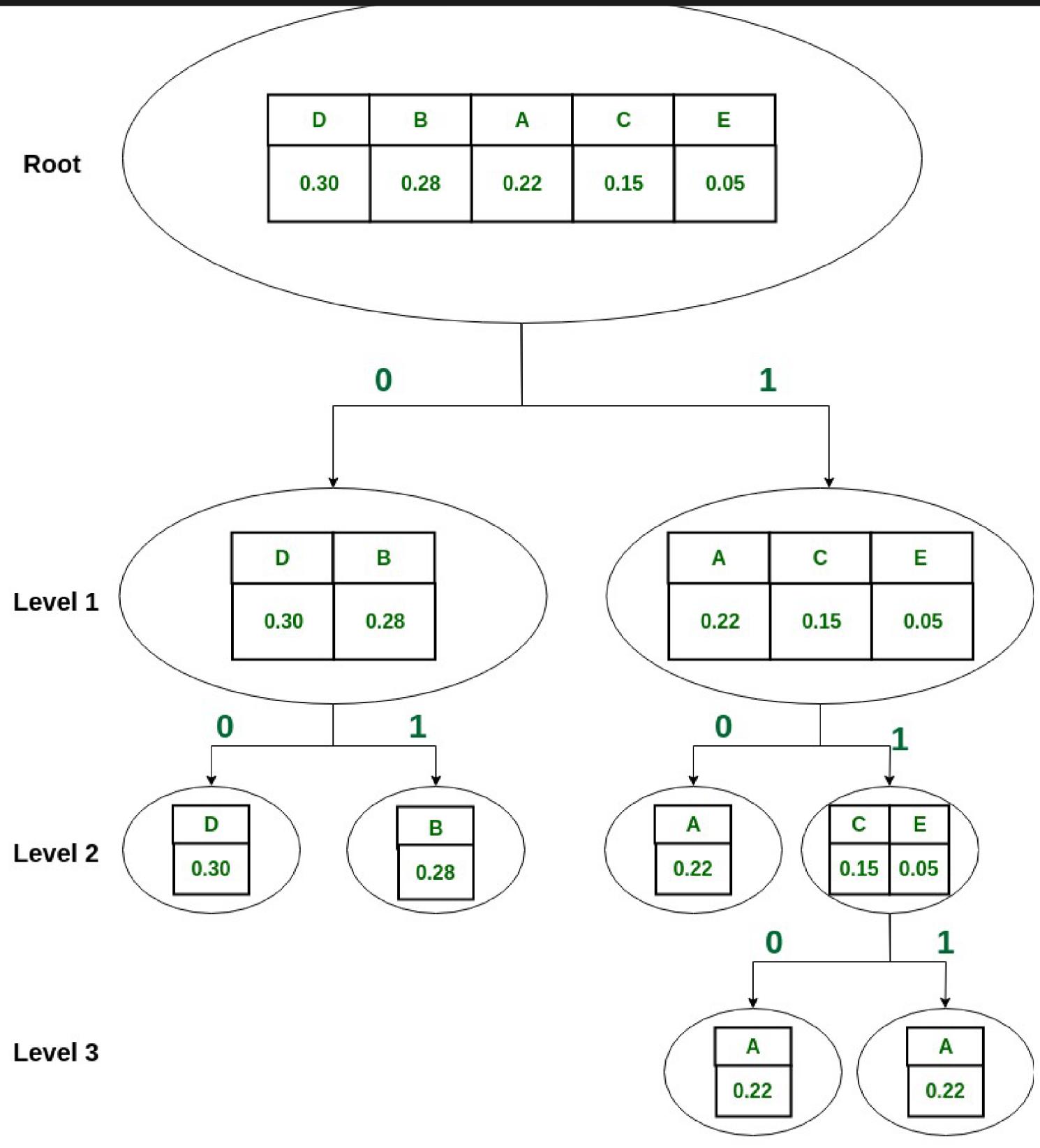
Steps of the Algorithm

- Calculate the probability of each symbol in the message.
- Sort the symbols in decreasing order of probability.
- Divide the set of symbols into two subsets such that the sum of the probabilities of the symbols in each subset is as close as possible to half the total probability.

SYMBOL	A	B	C	D	E
PROBABILITY OR FRQUENCY	0.22	0.28	0.15	0.30	0.05

THE SYMBOLS AND THEIR PROBABILITY / FREQUENCY ARE TAKEN AS INPUTS.

(In case of Frequency, the values can be any number)



- Assign a '0' to all symbols in the first subset and a '1' to all symbols in the second subset.
- Repeat steps 3-4 recursively for each group until each group contains only one symbol.
- Repeat steps 2-3 recursively on each subset until each branch contains only one character.
- Concatenate the codes assigned to each symbol to form the compressed message.

Purpose of the Algorithm

- The main purpose of the Shannon-Fano algorithm is to create a prefix code, where no code word is a prefix of any other code word, for a set of symbols with associated probabilities of occurrence.
- It encodes a message in a way that requires the least amount of space possible, while assigning variable-length codes to symbols, such that the most frequent symbols are assigned shorter codes, and the least frequent symbols are assigned longer codes.





ADVANTAGES OF THE ALGORITHM

- No information is lost during compression
- It is effective in compressing data with high redundancy, resulting in significant reduction in file size.
- It is simple to understand, making it suitable to use where computational resources are limited.

DISADVANTAGES OF THE ALGORITHM

- it may not be as effective in compressing data with low redundancy.
- It requires the probability of each input symbol, which may not be accurate.
- It compresses data without losing any information so it's not suitable for data such as in image or video.

Advantages and Disadvantages

Complexity

Time and Space complexity of the Algorithm



$O(n \log n)$

The TIME COMPLEXITY is $O(n \log n)$, where n is the number of characters. This measures the time to sort each symbol.



$O(n)$

The SPACE COMPLEXITY is $O(n)$, as it requires storing the probability and code for each symbol.

...



Thank you!

