

El algoritmo de la entropia: el mundo es ancho y ajeno

Yulissa Nataly Calcina Mango

UNIVERSIDAD NACIONAL DEL ALTIPLANO

Abril 22, 2024

1 INTRODUCCIÓN

2 FORMULA

3 CODIGO

4 IMAGEN

- La entropía, en el contexto de la teoría de la información, es una medida de incertidumbre o información promedio contenida en una fuente de información. Cuanto mayor sea la entropía, mayor será la incertidumbre o la imprevisibilidad en los datos provenientes de la fuente.
- El algoritmo de la entropía se utiliza para calcular la entropía de una fuente de información. En particular, el algoritmo de entropía de Shannon, propuesto por Claude Shannon, es el método más comúnmente utilizado para calcular la entropía. La fórmula de la entropía de Shannon es:



$$H(X) = - \sum_{i=1}^n P(x_i) \log_2 P(x_i)$$

Figure: formula de shannon

- Donde:
- $H(X)$ es la entropía de la fuente de información X .
- $P(x_i)$ es la probabilidad de que ocurra el simbolo x_i .
- El sumatorio se realiza sobre todos los simbolos x_i en el alfabeto de la fuente de información.

```
#include <iostream>
#include <fstream>
#include <string>
#include <unordered_map>
#include <cmath>

double calcularEntropiaShannon(const std::string& texto) {
    std::unordered_map<std::string, int> frecuenciaPalabras;
    int totalPalabras = 0;

    std::string palabraActual = "";
    for (char c : texto) {
        if (std::isalpha(c)) {
            palabraActual += std::tolower(c);
        } else if (!palabraActual.empty()) {
            frecuenciaPalabras[palabraActual]++;
            totalPalabras++;
            palabraActual = "";
        }
    }

    double entropia = 0.0;
```

```

for (const auto& par : frecuenciaPalabras) {
    double probabilidad = static_cast<double>(par.second) / totalPalabras;
    entropia -= probabilidad * std::log2(probabilidad);
}

return entropia;
}

int main() {
    std::ifstream archivo("el mundo es ancho y ajeno.txt");
    if (archivo.is_open()) {
        std::string texto((std::istreambuf_iterator<char>(archivo)), std::istreambuf_iterator<char>());
        archivo.close();

        double entropia = calcularEntropiaShannon(texto);
        std::cout << "Entropia por palabra: " << entropia << std::endl;
    } else {
        std::cerr << "Error al abrir el archivo." << std::endl;
    }

    return 0;
}

```

```

1  #include <iostream>
2  #include <fstream>
3  #include <string>
4  #include <unordered_map>
5  #include <cmath>
6
7  double calcularEntropiaShannon(const std::string& texto) {
8      std::unordered_map<std::string, int> frecuenciaPalabras;
9      int totalPalabras = 0;
10
11      std::string palabraActual = "";
12      for (char c : texto) {
13          if (std::isalpha(c)) {
14              palabraActual += std::tolower(c);
15          } else if (!palabraActual.empty()) {
16              frecuenciaPalabras[palabraActual]++;
17              totalPalabras++;
18              palabraActual = "";
19          }
20      }
21
22      double entropia = 0.0;
23      for (const auto& par : frecuenciaPalabras) {
24          double probabilidad = static_cast<double>(par.second) / totalPalabras;
25          entropia -= probabilidad * std::log2(probabilidad);
26      }
27
28      return entropia;
29  }
30
31  int main() {
32      std::ifstream archivo("el mundo es ancho y ajeno.txt");
33      if (archivo.is_open()) {
34          std::string texto((std::istreambuf_iterator<char>(archivo)), std::istreambuf_iterator<char>());
35          archivo.close();
36
37          double entropia = calcularEntropiaShannon(texto);
38          std::cout << "Entropia por palabra: " << entropia << std::endl;
39      } else {
40          std::cerr << "Error al abrir el archivo." << std::endl;
41      }
42
43      return 0;
44  }

```

C:\Users\ASUS\Documents\ENTROPIA.exe

Entropia por palabra: 7.61229

Process exited after 0.3439 seconds with return value 0
Presione una tecla para continuar . . .

sources Compile Log Debug Find Results Console Close

Compilation results...

```

-----
- Errors: 0
- Warnings: 0
- Output Filename: C:\Users\ASUS\Documents\ENTROPIA.exe
- Output Size: 3.10656261444092 MIB
- Compilation Time: 5.55s

```

Respuesta : la entropía por palabra del libro el mundo es ancho y ajeno es de :**7.61229**

paso a paso detallado sobre cómo calcular la entropía de un texto del libro de el mundo es ancho y ajeno en c++:

1.Preparación del texto:

- Asegúrate de tener el texto del libro "El mundo es ancho y ajeno" en un archivo de texto llamado "el mundo es ancho y ajeno.txt"

2. Bibliotecas utilizadas:

- `iostream`: Se utiliza para la entrada y salida estándar.
- `fstream`: Se utiliza para la lectura de archivos.
- `string`: Se utiliza para trabajar con cadenas de texto.
- `unordered_map`: Se utiliza para almacenar las frecuencias de las palabras.
- `cmath`: Se utiliza para operaciones matemáticas, en particular, para calcular el logaritmo base 2.

3. Función calcular Entropia Shannon:

- Esta función toma una cadena de texto como entrada y devuelve la entropía de Shannon calculada para ese texto.
- Se declara un `std::unordered_map` llamado `frecuenciaPalabras` para almacenar la frecuencia de cada palabra en el texto.
- Se inicializa una variable `totalPalabras` para contar el número total de palabras en el texto.
- Se recorre el texto carácter por carácter en un bucle `for`.
- Para cada carácter, se verifica si es una letra utilizando la función `std::isalpha()`. Si lo es, se convierte a minúsculas utilizando `std::tolower()` y se agrega a la palabra actual.

- Cuando se encuentra un carácter que no es una letra, se verifica si la palabra actual no está vacía. En ese caso, se incrementa su frecuencia en el mapa frecuencia Palabras, se actualiza el contador total Palabras y se reinicia la variable palabra Actual.
- Después de procesar todo el texto, se calcula la probabilidad de cada palabra dividiendo su frecuencia por el total de palabras.
- Se utiliza la fórmula de la entropía de Shannon para calcular la entropía, sumando el producto de la probabilidad de cada palabra por el logaritmo base 2 de esa probabilidad.
- La entropía calculada se devuelve como resultado de la función.

4. Función main:

- Se abre el archivo "el mundo es ancho y ajeno.txt" en modo de lectura utilizando `std::ifstream`.
- Si el archivo se abre correctamente, su contenido se lee en una cadena de texto `texto` utilizando un constructor que toma dos iteradores de flujo como argumentos. Esto lee todo el contenido del archivo en la cadena de texto.
- Se llama a la función `calcularEntropiaShannon` pasando el texto como argumento, y se almacena el resultado en la variable `entropia`.
- Finalmente, se imprime la entropía calculada por palabra utilizando `std::cout`.

5. Salida del programa:

- Si el archivo no se puede abrir, se imprime un mensaje de error en `std::cerr`.

Este código calcula la entropía de Shannon por palabra en un archivo de texto dado, proporcionando una medida de la incertidumbre promedio asociada con cada palabra en el texto.