

# Subsecuencia común más larga

Daniel Alessandro Aguilar Chombo

Junio 2020

## 1 Introducción

En este documento se abordara la definición del problema de *subsecuencia común más larga*, su complejidad, diferencias con otros problemas similares y sus aplicaciones en la vida real. Así mismo se presentara una solución utilizando programación dinamica.

## 2 Subsecuencia común más larga

El problema de *subsecuencia común más larga* se enfoca principalmente en encontrar, dado un conjunto de secuencias, la subsecuencia mas larga, que es común en dicho conjunto. De una manera formal podemos definirlo como, dadas dos secuencias  $X = \{x_1, \dots, x_n\}$  y  $Y = \{y_1, \dots, y_m\}$ , una subsecuencia  $Z = \{z_1, \dots, z_o\}$  se dice común si es subsecuencia simultaneamente de  $X$  y  $Y$ .

### 2.1 Complejidad

En casos generales donde se encuentra un número arbitrario de secuencias como entrada el problema es considera NP-Hard, cuando el número de secuencias es constante, el problema es resuelto en un tiempo polinomial, utilizando programación dinamica. Para el caso donde solo se utilizan dos secuencias como entrada, en tiempo puede ser expresado como  $O(n \times m)$ , donde  $n$  y  $m$  son las longitudes de las secuencias, en caso de que dichas longitudes sean iguales se puede expresar como  $O(n^2)$ .

### 2.2 Diferencias con otros problemas

La principal diferencia que cabe recalcar es con el problema de es con el problema de *secuencia común mas larga*, donde debemos recalcar la diferencia entre secuencia y subsecuencia, dicha diferencia yace en que la secuencia se compone de caracteres seguidos, mientras que en la subsecuencia no necesariamente debe de ser así.

## 2.3 Aplicaciones

Las aplicaciones del problema son muy variadas, dentro de las cuales podemos encontrar:

- Bioinformatica,
- Utilidades para la comparación de archivos
- Sistemas de versionado
- Linguistica computacional

### 2.3.1 Biological Applications, Of Mice And Men

Dentro de las ya mencionadas aplicaciones, resalta de manera muy particular las aplicaciones en la bioinformatica, esto lo podemos ver reflejado en ambas lecturas, las cuales remarcen las utilidades de este problema dentro de dicho campo, dentro de las cuales podemos encontrar:

- Secuenciación de ADN,
- Comparación de ADN
- Reconstrucción historica de ADN

## 3 Solución

Dentro de las soluciones que se pueden aplicar al problema se contemplan dos principalmente, mediante un algoritmo voraz o con base en la programación dinamica. Para que el problema de la subsecuencia común más larga sea optimo, se debe de optar por una solución basada en la programación dinamica.

Para poder diseñar un algoritmo basado en programación dinamica, se necesita encontrar una formula que cumpla con la con el principio de optimalidad, apartir del analisis podemos describir la formula de la siguiente manera.

$$f(x_i, y_j) = \begin{cases} 0, & \text{for } i = 0 \text{ or } j = 0 \\ f(x_i - 1, y_j - 1) + 1, & \text{for } x_i = y_j \\ \max\{f(x_i - 1, y_j), f(x_i, y_j - 1)\} & \text{for } x_i \neq y_j \end{cases}$$

Lo cual nos permite llegar a la siguiente implementación

---

**Algorithm 1:** LongestCommonSubsequence

---

```
Function longestCommonSubsequence(X: string, Y: string): int is  
  Data: X: string, Y: string, Z: int[ ], m: Length(X), n: Length(Y)  
  Result: Longitud de la subsecuencia común más larga  
  for i  $\leftarrow$  0 to m do  
    | Z  $\leftarrow$  Z [i, 0]  
  end  
  for j  $\leftarrow$  0 to n do  
    | Z  $\leftarrow$  Z [0, j]  
  end  
  for i  $\leftarrow$  0 to m do  
    for j  $\leftarrow$  0 to n do  
      if X [i] == Y [j] then  
        | Z [i, j]  $\leftarrow$  Z [i - 1, j - 1] + 1  
      else if Z [i - 1, j]  $\geq$  Z [i] [j - 1] then  
        | Z [i, j]  $\leftarrow$  Z [i - 1, j]  
      else  
        | Z [i, j]  $\leftarrow$  Z [i] [j - 1]  
      end  
    end  
  end  
  return Z [m, n]  
end
```

---

Dicha implementación puede ser encontrada en GitHub

## 4 Conclusión

En conclusión, el problema de la subsecuencia común más larga es de gran utilidad en diversos campos disciplinarios, permitiéndonos no solo comparar archivos, si no comparar, secuenciar y recuperar cadenas de ADN.