

Distancias de Secuencias

Daniel Bustos

May 26, 2024

10.

Se define la distancia entre dos secuencias de naturales $X = x_1, \dots, x_k$ e $Y = y_1, \dots, y_k$ como

$$d(X, Y) = \sum_{i=1}^k |x_i - y_i|.$$

Dado un conjunto de secuencias X_1, \dots, X_n , cada una de tamaño k , su grafo asociado G tiene un vértice v_i por cada $1 \leq i \leq n$ y una arista $v_i v_j$ de peso $d(X_i, X_j)$ para cada $1 \leq i < j \leq n$. Proponer un algoritmo de complejidad $O(kn^2)$ que, dado un conjunto de secuencias, encuentre el árbol generador mínimo de su grafo asociado.

Idea del algoritmo:

Vamos a construir el grafo en el cual todos están conectados con todos, y el peso de cada arista (X, Y) es $d(X, Y)$. ¿Cuánto nos cuesta construirlo? Dado que computar $d(X, Y)$ cuesta $O(k)$ y hay $\binom{n}{2}$ aristas, el costo total de construcción del grafo es

$$k \cdot \binom{n}{2} \in O(n^2 k).$$

Luego utilizamos el algoritmo de Prim con costo de $O(n^2)$ para generar el árbol mínimo, dejándonos un costo total de $O(n^2 k + n^2)$, que se incluye en $O(n^2 k)$.