École polytechnique — Majeure informatique

Promotion 2001

Cours de conception et analyse d'algorithmes

PC 7 – algorithmes et bio-informatique

1. Alignement de familles de séquences

i. Proposer un algorithme en $O(n^3)$ pour trouver le meilleur alignement de 3 séquences de longueur n pour une fonction de distance entre caractères $\delta : (\Sigma \cup \{-\})^2 \rightarrow \mathcal{N}$ donnée. δ mesure la distance entre les caractères, ainsi $\delta(\sigma, \sigma') = \delta(\sigma', \sigma) \geq 0$ et $\delta(\sigma, \sigma) = 0$.

ii. Proposer un algorithme qui étant donné deux chaînes de caractères A et B et une fonction de similarité δ, détermine en $O(n^2)$ les sous-chaînes α de A et β de B de valeur d'alignement maximum. Ici la fonction δ représente la valeur de l'alignement, ainsi $\delta(-,-) = 0$, $\delta(\sigma, \sigma) > 0$, et $\delta(\sigma, \tau) < 0$ pour $\sigma \neq \tau$.

2. Recherche de motif dans un texte

i. Proposer un algorithme de recherche d'un petit motif dans un texte, quand le motif a le droit de comporter, non seulement des lettres de l'alphabet Σ, mais aussi un symbole spécial “ ?" qui matche n'importe quel caractère.

ii. Proposer un algorithme de recherche d'un petit motif dans un texte, quand on a le droit d'avoir des erreurs ; on veut trouver un facteur du texte qui oubien est identique au motif oubien diffère du motif en seulement une lettre, s'il y en a un.

3. Sous-séquences maximales

i. Résoudre le problème des sous-séquences maximales dans le cas particuliers où pour tout i, on a $X_i \in \{-1, 1\}$.

ii. Généraliser le problème des sous-séquences maximales au cas de séquences circulaires.

iii.Généraliser le problème des sous-séquences maximales au cas où les variables X_i sont sur les sommets d'un arbre et où on souhaite trouver le sous-arbre dont la somme des X_i soit maximale.

4. Reconstruction d'arbre d'évolution.

Solit $G(D)$ le graphe complet dont l'arête $\{i, j\}$ a poids $D(i, j)$. Soit MD l'arbre couvrant minimum de $G(D)$. Montrer que D est ultra-métrique si et seulement si pour toute paire de noeuds $\{i, j\}$, dans l'arbre MD l'arête de poids maximum sur le chemin entre i et j a poids $D(i, j)$.