Partionnement spatial: approche k-means clustering

BM Bui-Xuan

Soit $\mathcal{S}$ un ensemble de n points répartis dans un plan en 2D.

Problème du k-*means clustering*

Le problème du k-*means clustering* consiste à trouver une partition de $\mathcal{S}$ en k parties telle que la somme des distances de chaque point par rapport au barycentre de sa partie est minimisée. Les heuristiques résolvant ce problème cherchent à fournir une telle répartition spatiale, dont la somme des distances est la plus petite possible (sans garantir que c'est la longueur optimale).

Il s'agit de proposer une heuristique pour le problème du k-*means clustering*, avec $k = 5$ et $\mathcal{S}$ représenté par les points contenus dans le fichier test.

Objectif : obtenir le plus petit score possible.

Version de k-*means clustering* avec contraintes :

Pendant la crise économique, le problème du k-*means clustering* avec restriction budgétaire s'énonce de la façon suivante. Soient k un entier, $s_1, s_2, \ldots, s_k \in \mathcal{S}$ des points de $\mathcal{S}$ appelés membres fondateurs, et B un réel appelé budget. Le problème du k-*means clustering* de budget B consiste à trouver k sous-ensembles disjoints $S_1, S_2, \ldots, S_k$ de $\mathcal{S}$ tels que :

— pour tout $1 \leqslant i \leqslant k$, on a $s_i \in S_i$;

— pour tout $1 \leqslant i \leqslant k$, la somme des distances des points de S_i par rapport au barycentre de S_i est au plus B ;

— le nombre d'éléments de $S_1 \cup S_2 \cup \cdots \cup S_k$ est maximisé.

Il s'agit de proposer une heuristique pour le problème du k-*means clustering* de budget B, avec $k = 5$, $\mathcal{S}$ représenté par les points contenus dans le fichier test, $s_1, s_2, \ldots, s_5$ les cinq premiers points dans ce fichier, et $B = 10101$.

Objectif : obtenir le plus grand score possible.