

Projet résolution distribuée de problèmes

De nos jours la classification non supervisée (clustering en anglais) est devenue de plus en plus importante car elle est utilisée dans plusieurs domaines(l'industrie, le business, le traitement d'images . . .etc). C'est en fait le processus de regroupement des données en groupes. Cela se fait de telle sorte que les objets d'un même groupe aient une grande similarité par rapport aux objets des autres groupes

L'algorithme k-médoide est un algorithme de classification non supervisée qui vise à regrouper un ensemble de données de n objets en k clusters. Il est robuste au bruit et aux valeurs aberrantes. L'algorithme commence par représenter les centres des clusters par des objets réels. Ensuite, chaque objet restant sera attribué au cluster le plus proche. Ce traitement est répété jusqu'à la stagnation du processus.

Entrée :

1. K : Nombre de clusters.
2. D : ensemble contenant N objets

Sortie :

Un ensemble de K clusters

Méthode :

1. Choisir K objets arbitrairement à partir de l'ensemble D pour former les centres initiaux.
2. Répéter :
 - Assigner chaque objet au cluster auquel il est le plus similaire se basant sur la valeur de distance entre les objets.
 - Choisir aléatoirement un objet non représentatif (O_r).
 - Calculer le coût total " S " des échanges d'objets représentatifs (O_j) avec (O_r)
 - Si $S < 0$ alors échanger (O_j) avec (O_r) pour former un ensemble de K objets représentatifs

Jusqu'à ce qu'il n'y ait pas de changement.

L'algorithme donne de meilleurs résultats d'un point de vue de la qualité de classification, mais il énormément de temps de calcul. Cet inconvénient rend l'algorithme applicable uniquement pour les données de petite taille.

L'objectif du projet :

Puisque la résolution distribuée des problèmes est une méthode efficace pour résoudre des problèmes complexes, Il est demandé de développer une application distribuée pour résoudre le problème de clustering. La solution proposée divise les données en partitions, qui sont traitées en parallèle. Les résultats des partitions sont ensuite fusionnés.

Evaluation :

L'évaluation du projet se base sur les points suivants :

- Un rapport de projet à remettre avant le 22/05/2022.
- Une démonstration de l'application.