

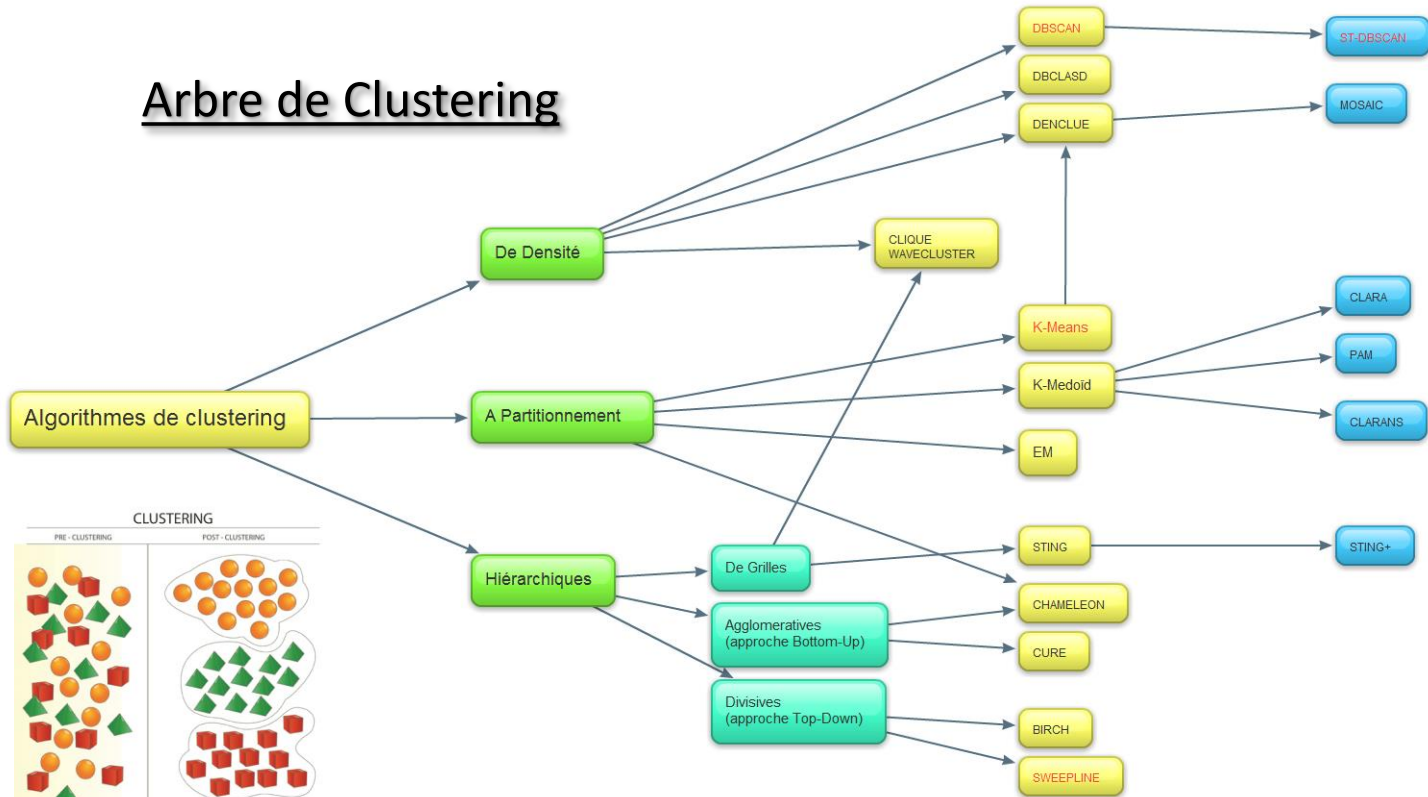
Algorithmes de Clustering Géographique

Etudiants: NADAT Mohamed et GAILLOT Antoine
Supervisé par VILLARD Lionel

Clustering ?

Le terme anglais **clustering**, qui se traduit en français par 'regroupement', désigne, en analyse statistique de données, les procédés mis en œuvre afin de classer les informations. Il fait référence aux méthodes de regroupement ou de partitionnement de données.

Arbre de Clustering



Les familles d'algorithmes

La méthode Hiérarchique se décompose en deux sous-méthodes, la méthode de regroupement (approche Bottom-Up) et la méthode de division (approche Top-Down). Le regroupement débute avec n clusters et à la fin de l'exécution de l'algorithme il ne reste plus qu'un seul cluster. Alors que la division commence avec un seul cluster pour aboutir à n clusters à la fin du traitement.

+
Traitement de large volume de données et insensibilité à l'ordre ni aux données inconsistantes (Grille).

-
Impossibilité de retour arrière pour certains algorithmes concernant la fusion (agglomératives) ou la scission (divisives).

La méthode de Partitionnement consiste à chercher les k meilleurs groupes d'un ensemble à n éléments. En effet, il cherche à former un certain nombre de clusters de données avec des caractéristiques communes.

+
Performant dans le traitement de données moyennement volumineuses.

-
Algorithmes sensibles aux données extrêmes exigeant un temps de traitement élevé.

La méthode de Densité consiste à faire des groupements selon la densité présente à l'intérieur d'un cluster, elle se base sur deux paramètres, le nombre de points (données) et la distance maximale acceptable autour d'un point (donnée) pour créer le cluster.

+
Traitement de large volume de données et haute dimensionnalité.

-
Nécessite une intervention humaine au préalable (algorithme supervisé), limité en terme de volume de données, consomme énormément de ressources.

*Cluster = Groupe

Références

- [1] Mamadou OUATTARA, *Fouille de données : vers une nouvelle approche intégrant de façon cohérente et transparente la composante spatiale*, 2010
- [2] Derya Birant & Alp Kut, *ST-DBSCAN: An Algorithm for clustering spatial-temporal data*, 2006
- [3] Krista Rizman Zalik & Borut Zalik, *A sweep-line algorithm for special clustering*, 2008
- [4] Khaled Alsabti & Sanjay Ranka & Vineet Singh, *An efficient k-means clustering algorithm*, 1997
- [5] Kiri Wagsta & Claire Cardie & Seth Rogers & Stefan Schroedl, *Constrained K-means Clustering with Background Knowledge*, 2001