

Approche biomimétique coopérative pour la visualisation de grands graphes multidimensionnels

Lydia Boudjeloud-Assala*, Hanane Azzag**

*Laboratoire d'Informatique Théorique et Appliquée EA 3097
Ile du Saulcy, F-57045 METZ CEDEX 1
lydia.boudjeloud@univ-metz.fr

**Laboratoire d'Informatique de Paris-Nord UMR CNRS 7030
99 Av. J.B. Clément, F-93430 VILLETANEUSE
hanane.azzag@lipn.univ-paris13.fr

Face à la quantité sans cesse grandissante de données stockées, les algorithmes de fouille et de visualisation de données doivent pouvoir être capable de traiter de grandes quantités de données. Une des solutions est d'effectuer un prétraitement des données permettant la réduction de la dimension des données sans perte significative d'informations. L'idée est donc de réduire l'ensemble de descripteurs avant de faire appel à la méthode de visualisation sous forme d'un graphe.

1 Approche coopérative

La solution proposée dans cet article est de faire coopérer deux méthodes biomimétiques pour la visualisation de graphes multidimensionnels. La première approche est un algorithme génétique pour la sélection de descripteurs (dimensions) pertinents développé par Boudjeloud-Assala (2005). L'algorithme génétique réduit l'espace de description des données en déterminant uniquement les descripteurs les plus pertinents. Chaque sous ensemble de descripteurs pertinents engendre un sous espace des données. L'évaluation des descripteurs se base sur une mesure de répartition des données Boudjeloud-Assala (2005), une fois un sous espace sélectionné le graphe de voisinages est construit par une autre approche biomimétique : algorithme utilisant des fourmis (agents) artificielles *AntGraph* (Lavergne, 2008). Nous nous intéressons principalement dans ce travail à la forme des graphes et non pas au résultat de la classification puisque celle ci dépend des résultats obtenues sur l'ensemble total des données. Nous comparons, donc, les visualisations obtenues d'une part sur l'ensemble total des descripteurs et d'autre part sur les différents sous ensembles de descripteurs obtenues par l'algorithme génétique. Nous remarquerons que dans certains cas (selon le sous espace sélectionné et les paramètres de la mesure d'évaluation) on se rapproche de la solution globale, et dans d'autres cas nous obtenons des sous espaces qui mettent en évidence des caractéristiques de données qui ne peuvent être vues dans l'espace total des données (des groupes plus homogènes, des données atypiques ...).