
Graphes de rigidité et structuration d'un système de classes

Christophe Osswald

Laboratoire E^3I^2 – EA 3876

ENSIETA

2, rue François Verny, 29206 Brest Cedex 9

L'un des objectifs de la classification est d'expliciter les relations entre éléments et classes. Les modèles usuels sont les hiérarchies, les pyramides et les systèmes de classes arborés, classiques en biologie. Ils nécessitent de faire le choix d'une structure pour approcher les données, et la plupart mènent à des problèmes NP-difficiles.

Dans le cadre de l'analyse de la similitude (Flament *et al.*, 1962-1981), il n'y a pas d'approximation, et l'interprétation des données passe par un *graphe de rigidité* le plus petit possible, tel que les classes d'un système donné en soient des classes connexes. Il existe plusieurs méthodes pour engendrer un système de classes à partir d'une dissimilarité : cliques maximales (classes naturelles), boules, 2-boules, réalisations (Brucker, 2003). Le problème est NP-difficile dans le cas général et pour les trois premières méthodes ; il existe un algorithme en $O(n^4)$ pour les réalisations.

Il est aisé d'identifier les hyperarbres comme étant les systèmes de classes rigides sur un arbre. Ceci s'étend aux *hypercycles*, dont un graphe de rigidité est un cycle, en $O(n^4)$ opérations, ainsi que les systèmes de classes dont un graphe de rigidité a n arêtes.

Pour une dissimilarité arboricole (dont les classes admettent un arbre de rigidité), les arbres de rigidité des boules sont aussi ceux des 2-boules, des classes et des réalisations. Au-delà de ce modèle, l'équivalence entre les graphes de rigidité de ces systèmes de classes se limite à une chaîne d'inclusion. Le nombre d'arêtes d'un graphe de rigidité minimum des réalisations d'une dissimilarité quelconque nous permet ainsi de construire une mesure de la structuration d'icelle. Nous confrontons cette mesure à plusieurs modèles de dissimilarité aléatoire, afin notamment d'analyser des dissimilarités réelles et graphiques.

Ces méthodes ont été appliquées à des données issues de la psychologie de la mémoire, de l'analyse de données textuelles, de l'imagerie sonar et de la génétique.

[BRU 79] FLAMENT C., DEGENNE A., VERGES P., “Analyse de similitude ordinale”, *Informatique et Sciences Humaines*, vol. 40-41, 1979, pp 223-231.

[FLA 62] FLAMENT C., “L'analyse de similitude”, *Cahiers du Centre de Recherche Opérationnelle*, vol. 4, 1962, pp 63-97.

[BRU 03] BRUCKER F., “Réalizations de dissimilarités”, *Rencontres de la Société Francophone de Classification*, 2004, pp 7-10.