

# Une étude d'algorithmes de classification supervisée basée sur les treillis de Galois

Huaiyu Fu, Huaiguo Fu, Patrick Njiwoua, Engelbert Mephu Nguifo  
CRIL-CNRS FRE2499, Université d'Artois - IUT de Lens  
Rue de l'université SP 16, 62307 Lens cedex. France  
{fu,mephu}@cril.univ-artois.fr

La classification supervisée est une des tâches de la fouille de données. C'est un processus comprenant deux phases : apprentissage et classement. La phase d'apprentissage consiste à construire un modèle (ou classifieur) qui décrit un ensemble prédéterminé de classes d'exemples. La phase de classement consiste à utiliser le modèle pour affecter une classe à un nouvel exemple.

Plusieurs méthodes de classification supervisée publiées dans la littérature s'appuient sur des techniques différentes [Cornuejols et Miclet, 2002]. Nous nous intéressons aux méthodes basées sur les treillis de concepts [Ganter et Wille, 1999]. Les treillis de concepts (ou treillis de Galois) sont une structure mathématique permettant de représenter les classes non disjointes sous-jacentes à un ensemble d'objets décrits à partir d'un ensemble d'attributs. Ces classes non disjointes sont appelées concepts formels.

Des méthodes de classification basées sur le treillis de concepts ont été développées et comparées à des méthodes standards, mais aucune comparaison n'a encore été faite entre ces méthodes. Nous avons décrit suivant un même schéma de principe, les systèmes publiés à notre connaissance dans la littérature, et présentant les deux phases de classification : GRAND, LEGAL, GALOIS, RULEARNER, CIBLE et CLNN & CLNB [Njiwoua et Mephu Nguifo, 2002].

Nous avons implémenté 4 de ces algorithmes (GRAND, LEGAL, GALOIS and RULEARNER) en JAVA, permettant ainsi pour une première fois de comparer ces méthodes entre-elles sur une même plate-forme. Les résultats expérimentaux montrent que ces systèmes ont des performances en terme de taux de précision meilleures que C4.5, sur certaines données qu'il conviendra de caractériser. Néanmoins les temps de calcul restant élevés, il est nécessaire de poursuivre l'effort de recherche de méthodologie d'élagage du treillis pour favoriser son usage pour la classification supervisée.

## Références

- [Cornuejols et Miclet, 2002] A. Cornuejols et L. Miclet. *Apprentissage Artificiel : Méthodologie et Applications*. Eyrolles, Paris, France, 2002. 630 pages.
- [Ganter et Wille, 1999] B. Ganter et R. Wille. *Formal Concept Analysis. Mathematical Foundations*. Springer, 1999.
- [Njiwoua et Mephu Nguifo, 2002] P. Njiwoua et E. Mephu Nguifo. Classification supervisée et treillis de galois. Technical report, CRIL, Université d'Artois, France, Novembre 2002.