

Combinaison de classification supervisée et non-supervisée par la théorie des fonctions de croyance

Fatma Karem*, Mounir Dhibi*
Arnaud Martin**

*Unité de Recherche PMI 09/UR/13-0
Campus Universitaire Zarouk Gafsa 2112, Tunisie
fatoumacy@yahoo.fr, mounir.dhibi@ensta-bretagne.fr,
**Université de Rennes 1, UMR 6074 IRISA
Rue Edouard Branly BP 30219, 22302 Lannion Cedex, France
Arnaud.Martin@univ-rennes1.fr

Résumé. Nous proposons dans cet article une nouvelle approche de classification fondée sur la théorie des fonctions de croyance. Cette méthode repose sur la fusion entre la classification supervisée et la classification non supervisée. En effet, nous sommes face à un problème de manque de données d'apprentissage pour des applications dont les résultats de classification supervisée et non supervisée sont très variables selon les classificateurs employés. Les résultats ainsi obtenus sont par conséquent considérés comme incertains.

Notre approche se propose de combiner les résultats des deux types de classification en exploitant leur complémentarité *via* la théorie des fonctions de croyance. Celle-ci permet de tenir compte de l'aspect d'incertitude et d'imprécision. Après avoir dresser les différentes étapes de notre nouveau schéma de classification, nous détaillons la fusion de classificateurs. Cette nouvelle approche est appliquée sur des données génériques, issues d'une vingtaine de bases de données. Les résultats obtenus ont montré l'efficacité de l'approche proposée.

1 Introduction

La classification est un moyen utile d'organisation et de hiérarchisation des données. Le but de la classification non-supervisée est de trouver des groupes compacts et bien séparés dans un ensemble de données et donc d'affecter à chaque observation une étiquette de classe qui matérialise l'appartenance de celle-ci aux classes dégagées. De plus on souhaite pouvoir également affecter à toute nouvelle observation une étiquette. Cette situation n'est pas rencontrée en classification supervisée puisque les observations disponibles sont déjà étiquetées le but à atteindre est d'affecter à une nouvelle observation, une classe préexistante et apprise sur les données d'apprentissage. Le problème dans le contexte de la classification non-supervisée est plus difficile puisqu'aucune information n'est fournie sur l'appartenance des données à telle ou telle classe. Cette appartenance est généralement déduite à partir de la répartition spatiale des points expliquée par Campedel (2005). Indépendamment du type de classification, le choix de