

Extraction de connaissances provenant de données multisources pour la caractérisation d'arythmies cardiaques

Elisa Fromont, René Quiniou, Marie-Odile Cordier

IRISA, Campus de Beaulieu F-35042 RENNES Cedex France

{efromont, quiniou, cordier}@irisa.fr,

<http://www.irisa.fr/dream>

Résumé. Nous nous intéressons au problème de l'apprentissage par programmation logique inductive de règles symboliques caractérisant des arythmies cardiaques à partir de données multisources hétérogènes telles que les différentes voies d'un électrocardiogramme ou la mesure de la pression artérielle. Une première approche consiste à agréger les données dans la base d'apprentissage puis à effectuer un apprentissage directement à partir de ces données transformées. Cette méthode d'apprentissage global est peu performante et difficile à mettre en œuvre quand le volume des données est important. Nous proposons une alternative plus efficace permettant de tirer profit d'apprentissages effectués préalablement sur chacune des sources indépendamment, pour construire automatiquement un biais permettant de restreindre l'espace de recherche lors d'un apprentissage multisource et ainsi traiter ces données complexes. Les résultats obtenus par les deux méthodes sont analysés et comparés. Ils montrent que la méthode proposée permet de gagner un ordre de grandeur sur les temps d'apprentissage.

1 Introduction

Nous nous intéressons à l'extraction de connaissances, par exemple sous forme de règles de classification, à partir de données provenant de plusieurs sources reflétant un même phénomène. Ces données peuvent être soit totalement indépendantes, soit corrélées. Elles peuvent aussi avoir une pertinence ou une fiabilité différente selon la source dont elles proviennent.

On peut alors se demander si une seule source contient suffisamment d'information pour traiter un problème donné ou s'il est préférable de profiter de la complémentarité des différentes sources. Dans le cas monosource, on acquiert de la connaissance à partir de chacune des sources indépendamment, puis on utilise une procédure de décision évaluant les résultats obtenus. Cette procédure de décision basée sur la fiabilité et la pertinence des sources [Cholvy, July 2003], est difficile à mettre en œuvre sans une connaissance précise de chacune des sources. Des méthodes telles que le vote [Dubois *et al.*, 2001], la combinaison de classifieurs [Wemmert et Gancarski, 2002], des règles de "bonne fusion" issues des connaissances sur le domaine [Bloch *et al.*, 2001], etc. peuvent alors être utilisées. Cependant, ces méthodes, même si elles sont effi-