

Utilisation des réseaux bayésiens dans le cadre de l'extraction de règles d'association

Clément Faure^{*,**}, Sylvie Delprat^{*}
Alain Mille^{***}, Jean-François Boulicaut^{**}

^{*}EADS CCR, Centreda 1, F-31700 Blagnac
{clement.faure, sylvie.delprat}@eads.net

^{**}LIRIS UMR 5205, INSA Lyon, Bâtiment Blaise Pascal,
F-69621 Villeurbanne cedex

^{***}LIRIS UMR 5205, Université Lyon 1, Nautibus,
F-69622 Villeurbanne cedex
{amille, jboulica}@liris.cnrs.fr

Résumé. Cet article aborde le problème de l'utilisation d'un modèle de connaissance dans un contexte de fouille de données. L'approche méthodologique proposée montre l'intérêt de la mise en œuvre de réseaux bayésiens couplée à l'extraction de règles d'association dites delta-fortes (membre gauche minimal, fréquence minimale et niveau de confiance contrôlé). La découverte de règles potentiellement utiles est alors facilitée par l'exploitation des connaissances décrites par l'expert et représentées dans le réseau bayésien. Cette approche est validée sur un cas d'application concernant la fouille de données d'interruptions opérationnelles dans l'industrie aéronautique.

1 Introduction

Un des objectifs de l'extraction de connaissances à partir de données consiste à fournir des énoncés valides et utiles aux utilisateurs propriétaires de ces données. L'utilité de ces énoncés est d'autant plus grande qu'ils décrivent une réalité du domaine non encore explicitée jusqu'ici, autrement dit, une nouvelle connaissance.

Nous nous intéressons à l'extraction de connaissances au moyen de règles descriptives comme les règles d'association (Agrawal et al., 1993). Les problèmes posés par l'extraction de telles règles ont été étudiés intensivement ces dix dernières années. Bien que l'extraction de toutes les règles fréquentes et valides soit difficile dans de grands jeux de données, des dizaines d'algorithmes efficaces ont été proposés (Goethals et Zaki, 2003, par exemple). Un second problème concerne le nombre considérable de règles qui peuvent être fréquentes et valides et donc extraites. Une première solution consiste à rechercher des couvertures des ensembles de règles, ou si l'on préfère, à éliminer des règles redondantes. Des travaux importants dans cette direction concernent l'exploitation de représentations condensées des ensembles fréquents comme les ensembles fermés (Pasquier et al., 1999; Boulicaut et al., 2000) ou bien les ensembles δ -libres (Boulicaut et al., 2003). (Jeu, 2002) est une étude assez complète de ces propositions.