

Extraction efficace de règles graduelles

Lisa Di Jorio*, Anne Laurent* Maguelonne Teisseire*

*LIRMM – Université de Montpellier 2 – CNRS
161 rue Ada, 34392 Montpellier – FRANCE
{dijorio, laurent, teisseire}@lirmm.fr
<http://www.lirmm.fr/~{dijorio, laurent, teisseir}>

Résumé. Les règles graduelles suscitent depuis quelques années un intérêt croissant. De telles règles, de la forme “*Plus (moins) A_1 et ... plus (moins) A_n alors plus (moins) B_1 et ... plus (moins) B_n* ” trouvent application dans de nombreux domaines tels que la bioinformatique, les contrôleurs flous, les relevés de capteurs ou encore les flots de données. Ces bases, souvent composées d’un grand nombre d’attributs, restent un verrou pour l’extraction automatique de connaissances, car elles rendent inefficaces les techniques de fouille habituelles (règles d’association, clustering...). Dans cet article, nous proposons un algorithme efficace d’extraction d’itemset graduels basé sur l’utilisation des treillis. Nous définissons formellement les notions de gradualité, ainsi que les algorithmes associés. Des expérimentations menées sur jeux de données synthétiques et réels montrent l’intérêt de notre méthode.

1 Introduction

L’évolution des capteurs, de plus en plus précis, robustes et abordables, permet l’acquisition de nombreuses mesures fiables, produisant ainsi des bases de données numériques denses et volumineuses. Cependant, même si la fouille de données quantitatives est un domaine étudié depuis plusieurs années (Srikant et Agrawal (1996)), la densité des bases pose de nouvelles problématiques, car elle rend inefficaces les techniques de fouille habituelles. Pourtant, les experts sont de plus en plus en attente de méthodes efficaces afin de prendre des décisions ou d’analyser différents comportements. Beaucoup de domaines sont concernés, comme par exemple le domaine biomédical, où les principales découvertes passent par l’analyse du génome (contenant plusieurs milliers de gènes, pour peu de patients), ou encore le domaine de l’analyse de capteurs et de flots de données où les comportements fréquents servent à la surveillance ou à la détection / prévention de pannes.

Nous nous intéressons à la découverte de connaissances au moyen de *règles graduelles*. Les règles graduelles modélisent des co-variations fréquentes sur les valeurs d’items, et sont de la forme “*plus (moins) A_1 et ... plus (moins) A_n , alors plus (moins) B_1 et ... plus (moins) B_p* ”. Ces règles suscitent depuis quelques années un intérêt croissant (Hüllermeier (2002); Berzal et al. (2007)). La notion de gradualité, et plus particulièrement de règles graduelles, a majoritairement été étudiée dans le domaine du flou. Celles-ci étaient utilisées dans le but de modéliser des systèmes experts. L’accent n’est alors pas mis sur la manière de les extraire, mais