

Classification préalable à la recherche de règles d'association

Cette communication décrit l'utilisation conjointe des méthodes de recherche de règles d'association et de classification de variables dans le cadre d'une application du secteur automobile où plus de 80000 véhicules sont décrits par plus de 3000 attributs binaires rares. Rechercher des règles d'association (Agrawal & Srikant, 1994) sur un grand ensemble d'événements rares conduit à une profusion de résultats difficiles à interpréter de part leur nombre et leur complexité. Nous proposons de réaliser une classification préalable des variables (Plasse et al., 2005) afin d'obtenir des groupes homogènes d'attributs puis de rechercher les règles d'association à l'intérieur de chacun de ces groupes.

Nous avons réalisé plusieurs classifications de variables (Nakache & Confais, 2005) à l'aide d'un algorithme divisif et d'un algorithme agglomératif avec différents indices de similarité entre variables. Après avoir comparé deux à deux toutes les partitions obtenues grâce à l'indice de Rand. Nous avons ensuite procédé à la recherche des règles d'association à l'intérieur de chaque classe de variables, dans toutes les partitions. La classification descendante permet directement une réduction importante du nombre de règles, contrairement aux classifications ascendantes, où la réduction est moindre. Cependant ces dernières permettent de mettre en évidence une forte liaison entre certains attributs dont les associations multiples produisent des règles complexes en grand nombre. En isolant ces attributs, la réduction du nombre de règles obtenues est alors supérieure à 99% quelque soit la méthode de classification employée. De plus, la complexité des règles n'excède pas 5 attributs.

Après avoir montré l'apport d'une classification de variables préalable à la recherche de règles d'association, nous travaillons actuellement sur l'apport de la classification croisée à cette approche.