Génération de règles par compression

Fabrice Muhlenbach — Djamel A. Zighed — Sophie d'Hondt

Laboratoire ERIC
Université Lumière Lyon 2
Bâtiment L - 5 avenue Pierre Mendès-France
69676 BRON Cedex - FRANCE
muhlenba@eric.univ-lyon2.fr, zighed@univ-lyon2.fr, sophie_dhondt@yahoo.fr

RÉSUMÉ. *Au sein des technologies d'extraction automatique de connaissances, les méthodes procédant par des graphes d'induction ont l'avantage de proposer des modèles compréhensibles par la production de règles logico-déductives. Une critique faite à ces méthodes est qu'elles produisent parfois trop de règles, ce qui a pour effet de ne plus rendre le modèle compréhensible. Nous proposons une nouvelle méthode de génération des règles de production – appelée « Data Squeezer » – qui se différencie des autres méthodes par sa démarche ascendante. En procédant par la compression des données, cette méthode permet d'obtenir un nombre de règles réduit, tout en gardant de bonnes capacités d'apprentissage et de généralisation, et de résoudre des problèmes d'apprentissage tels que le OU exclusif.*

ABSTRACT. *Within data mining technologies graph induction methods give the benefit of producing models that are understandable by rules. However these methods sometimes produce too many rules that make lose this comprehensibility. We propose a new method for generating production rules called "Data Squeezer", which is bottom-up unlike other methods. By squeezing the data, this method produces a reduced number of rules while having good results for classification in learning and generalisation. This approach can also resolve some machine learning problems like the XOR.*

MOTS-CLÉS : *Apprentissage automatique, ECD, Modèle prédictif, Règles de production.*

KEY WORDS: *Machine learning, Data mining, Predictive model, Production rules.*