

Chapitre 6 : Analyse statistique implicative entre variables vectorielles

Régis Gras*, Raphaël Couturier**

*Equipe Connaissances et Décision, Laboratoire d'Informatique de Nantes Atlantique
Ecole Polytechnique de l'Université de Nantes, UMR 6241
La Chantrerie BP 60601 44306 Nantes cedex
regisgra@club-internet.fr

** Institut Universitaire de Technologie de Belfort,
BP 527, rue E. Gros, 90016 Belfort cedex
raphael.couturier@iut-bm.univ.fcomte.fr

Résumé. Nous nous plaçons ici dans le cadre de la méthode d'analyse de données, l'analyse statistique implicative (A.S.I.). A l'instar de ce que nous avons fait pour passer des variables binaires aux variables numériques ou aux variables-intervalles, nous étendons le champ des traitements aux variables à valeurs vectorielles. Nous établissons un indice permettant de mesurer la qualité d'une règle entre variables vectorielles. Nous traitons des exemples portant l'un sur le baccalauréat, l'autre sur l'examen des critères de convergence des économies de l'Union Européenne.

1 Introduction

Les développements théoriques de l'analyse de données offrent des retombées enrichissantes pour l'Extraction de Connaissances. Ses développements et sa vitalité ne sont pas étrangères aux échanges induits. Par exemple, la construction d'indices permettant d'affecter une mesure non symétrique à des règles d'inférence partielle fournit des points d'application à l'extraction et à la représentation de règles d'association imprécises entre attributs binaires décrivant une population. Les démarches fondamentales se ramènent à la prise en compte d'une problématique commune aux deux domaines ; il s'agit de découvrir et de quantifier des règles non symétriques pour modéliser des relations du type "*si a, alors presque b*". Qu'il s'agisse de réseaux bayésiens (Pearl, 1988), de treillis de Galois (Bernard et Poitrenaud, 1999) ou de fouille de règles (Agrawal et al, 1993) de très nombreuses mesures ont été proposées pour quantifier la pertinence de ces quasi-implications et optimiser leur extraction (par ex. (Hilderman et Hamilton, 1999) ou (Tan et Kumar, 2000)), Des travaux sur la qualité des règles d'association ont permis comparer leurs mesures selon des points de vue subjectifs et objectifs (Lenca et al, 2004). Cependant, à notre connaissance, ces travaux se limitent généralement à l'étude de mesures pour des règles entre attributs binaires ou conjonction de tels attributs.

Or les situations réelles, y compris celles pour lesquelles l'analyse implicative a créé son modèle statistique (la didactique des mathématiques), conduisent au traitement d'autres types