

Typicalité et contribution des sujets et des variables supplémentaires en Analyse Statistique Implicative

Régis Gras *, Jérôme David*, Jean-Claude Régnier**, Fabrice Guillet*

* LINA– Ecole Polytechnique de l'Université de Nantes

La Chantrerie BP 60601 44306 Nantes cedex

regisgra@club-internet.fr, jerome.david, fabrice.guillet@polytech.univ-nantes.fr

<http://www.sciences.univ-nantes.fr/lina/>

**EA 3727 Savoirs, Diversité et Professionnalisation, Lyon 2

86, rue Pasteur 69365 Lyon cedex 07

Jean-claude.regnier@univ-lyon2.fr

Résumé. L'analyse statistique implicative traite des tableaux sujets \times variables afin d'extraire règles et métarègles statistiques entre les variables. L'article interroge les structures obtenues représentées par graphe et hiérarchie orientés afin de dégager la responsabilité des sujets ou des groupes de sujets (variables supplémentaires) dans la constitution des chemins du graphe ou des classes de la hiérarchie. On distingue les concepts de typicalité pour signifier la proximité des sujets avec le comportement moyen de la population envers les règles statistiques extraites, puis de contribution pour quantifier le rôle qu'auraient les sujets par rapport aux règles strictes associées. Un exemple de données réelles, traité à l'aide du logiciel CHIC, illustre et montre l'intérêt de ces deux concepts.

1 Introduction

Les données traitées par l'analyse statistique implicative (en abrégé : A.S.I.) se présentent sous forme de tableaux numériques croisant une population E de sujets, ou individus ou objets, associé chacun à une ligne, et un ensemble V de variables simples ou conjointes (attributs binaires, variables numériques, rang, intervalle) chacune associée à une colonne. A l'intersection de la ligne x et de la colonne j figure la valeur prise par le sujet x selon la variable j . La finalité première de l'A.S.I. vise à dégager de V ou de l'ensemble de toutes les conjonctions d'éléments de V ¹, des règles d'association non symétrique, contrairement à la similarité, sur une base statistique, du type : « si la variable ou une conjonction de variables a est observée sur E alors la variable b a tendance à être observée », règle notée $a \Rightarrow b$. Une mesure de qualité, non symétrique, de telles règles² est définie par :

¹ Dorénavant nous continuerons à noter V , pour éviter des notations excessives, aussi bien l'ensemble des variables que celui de toutes les conjonctions de ses éléments.

² D'autres mesures existent comme celle d'(Agrawal et al.,1993) basée sur les deux paramètres : support (fréquence de a et b) et confiance (fréquence conditionnelle de b sachant a)