

Chapitre 2 : Qualité d'un graphe implicatif : variance implicative

Régis Gras*. Jean-Claude Régnier**

*LINA– Ecole Polytechnique de l'Université de Nantes, UMR 6241
La Chantrerie BP 60601 44306 Nantes cedex

regisgra@club-internet.fr

** Université de Lyon - UMR 5191 ICAR
ENS-LSH 15, Parvis René Descartes BP 7000 69342 LYON cedex 07
jean-claude.regnier@univ-lyon2.fr

Résumé. Un graphe pondéré, sans cycle, constitue une des représentations d'un ensemble de règles d'association implicative extraites d'un tableau numérique croisant variables et sujets. Le problème de son homogénéité, de sa cohérence et donc de la pertinence des interprétations de l'expert se pose dès lors qu'en Analyse Statistique Implicative (A.S.I.) il est possible de faire varier le seuil de représentation des règles partielles. Nous présentons ici le concept de variance implicative à l'instar du concept classique de variance afin de qualifier l'homogénéité de la représentation. Elle s'appuie sur une métaphore de répulsion vs consistance implicatives mutuelles entre deux variables binaires à partir de leur différence symétrique.

1 Introduction

Le texte qui suit, tente de répondre à une question posée récemment par Michel Oris et Gilbert Ritschard, (2007), question qui se ramène à celle-ci : dans quelle mesure peut-on affirmer qu'un graphe implicatif présente une « bonne » qualité de structure arborescente par rapport aux données ? Pourrait-on définir une mesure de type inertiel, par exemple, comparable à celle qui permet de qualifier une partition en classes homogènes et « convenablement séparées » ? ¹

Dans la Partie 1, chapitre 4, nous avons établi un critère probabiliste permettant de qualifier des niveaux de la hiérarchie orientée de R-règles, puis celle de la hiérarchie entière. Dans l'ouvrage *L'implication statistique. Nouvelle méthode exploratoire de données* (Gras et al, 1996), nous avons défini un critère numérique, « La variance statistique de classes cohésitives », pour quantifier la qualité d'une hiérarchie de R-règles à ses différents niveaux

¹ Je cite : « ...limites de l'ASI....absence de critère permettant de juger de la pertinence statistique globale du modèle retenu. Quel pourrait être un équivalent de la déviance utilisée en modélisation statistique ou de la part d'inertie reproduite en analyse factorielle ? » (M.Oris et G. Ritschard, dans « Dynamique professionnelle dans la Genève du 19ème, enseignements d'une analyse de statistique implicative", Actes de ASI 4,, octobre 2007)