

Quelques critères pour une mesure de qualité de règles d'association

Régis Gras*, Raphaël Couturier**, Julien Blanchard*
Henri Briand*, Pascale Kuntz*, Philippe Peter*

*École polytechnique de l'université de Nantes
regis.gras@club-internet.fr

{julien.blanchard,henri.briand,pascale.kuntz,philippe.peter}@polytech.univ-nantes.fr

**Laboratoire d'Informatique de Franche Comté (LIFC)
couturie@iut-bm.univ-fcomte.fr

Résumé. De nombreuses mesures de qualité d'une règle d'association implicative existent mais peu d'entre elles font état des raisons qui ont guidé leurs choix épistémologiques. Nous présentons ici quelques critères qui permettraient de donner un sens aux indices qui servent à quantifier la qualité de l'association. Ensuite, nous développons, en les justifiant, les différents choix que nous avons faits pour mesurer la qualité des implications entre variables de nature binaire ou non binaire, ainsi qu'entre classes ordonnées de variables. Nous confrontons ces choix aux critères énoncés préalablement. Quelques simulations permettent d'illustrer graphiquement la différence de comportement entre certains indices classiques au regard de ces critères.

1 Introduction

Les connaissances *opératoires* et *décisionnelles* de l'homme sont principalement constituées et activées selon deux composantes : celles des faits et celles des règles entre les faits ou, à un niveau supérieur, entre des règles elles-mêmes. Ce sont ses apprentissages à travers son éducation, sa culture ou ses expériences personnelles qui lui permettent une élaboration progressive de ces connaissances, même si des régressions, des remises en cause viennent continûment lui assurer un équilibre fonctionnel. Extraire des connaissances à partir des données, consiste donc à effectuer un traitement cognitif, psychologique ou automatisé, à partir d'un ensemble de faits afin de mettre en évidence principalement des associations entre faits, mais également, « cerise sur le gâteau », des concepts et des structures. Parmi les associations, les règles de concomitance (« a et b apparaissent en même temps ») qui sont essentiellement symétriques, au même titre que la similarité, diffèrent alors des règles implicatives (« si a alors b ») qui sont dissymétriques, au même titre que l'implication (non nécessairement causale). Or les règles, quelles qu'elles soient, se constituent de façon stable dès lors que le nombre de succès quant à leur qualité prédictive ou anticipatrice atteint un certain niveau de confiance en deçà duquel elles seront susceptibles d'être plus imprudemment mises en oeuvre. En revanche, ce niveau étant atteint ou dépassé, l'économie individuelle le fera résister à son abandon ou, peut-être, à sa simple critique. En effet, il est coûteux,