

Chapitre 11 : Historique et fonctionnalités de CHIC

Raphaël Couturier* et Saddo Ag Almouloud**

* Laboratoire d'Informatique de l'université de Franche-Comté (LIFC),
IUT de Belfort-Montbéliard, BP 527, 90016 Belfort, France

raphael.couturier@iut-bm.univ-fcomte.fr

** Pontifícia Universidade Católica de São Paulo - PUC/SP
Rua Marquês de Paranaguá, 111, Consolação, São Paulo - SP - Brasil
saddoag@gmail.com

<http://www.pucsp.br/pos/edmat>

Résumé. CHIC permet d'utiliser la plupart des méthodes définies dans le cadre de l'ASI (Implication Statistique Implicative). Il a pour objectif de découvrir les implications les plus pertinentes entre les variables d'un ensemble de données. Pour cela, il propose d'organiser les implications sous forme d'une hiérarchie cohésive (orientée) ou un graphe implicatif. De plus, il permet d'obtenir une hiérarchie des similarités (non orientée) basée sur les ressemblances des variables. Ce papier décrit l'histoire, les caractéristiques et l'usage de CHIC.

1 Introduction

L'analyse statistique implicative (ASI) a été développée par Régis Gras et ses collaborateurs. Elle permet d'établir des règles d'association à partir d'un ensemble de données croisant sujets et variables. Le but initial de cette méthode a été de répondre à la question : "Si un objet possède une propriété, est ce qu'il en possède une autre ?". Bien entendu lorsque la réponse est totalement affirmative, il est facile de répondre à cette question. Cependant, il est possible que ce ne soit pas le cas, alors on peut constater que des tendances apparaissent. L'ASI a pour objectif de mettre en évidence de telles tendances dans un ensemble de propriétés. Comparée aux autres méthodes statistiques qui permettent de générer des règles d'association, l'ASI se distingue par le fait qu'elle utilise une mesure non linéaire qui satisfait des critères importants. Tout d'abord, cette mesure est basée sur l'intensité d'implication qui mesure le degré de surprise inhérent à une règle. Ainsi, les règles triviales qui sont potentiellement évidentes et connues de l'expert sont supprimées. Cette intensité d'implication peut être renforcée par le degré de validité, défini par rapport à l'entropie de Shannon, si l'utilisateur choisit ce mode de calcul. Dans ce cas, la mesure ne prend pas simplement en compte la validité de la règle, mais aussi sa contraposée. En effet, quand une règle d'association est estimée valide, c'est-à-dire que l'ensemble des items A est fortement proche de l'ensemble des items B , alors il est légitime et intuitif d'attendre que la contraposée soit valide, c'est-à-dire que l'ensemble des items non- B soit fortement