

# Notion de sémantiques bien-formées pour les règles

Marie Agier<sup>\*\*\*</sup>, Jean-Marc Petit<sup>\*\*</sup>

<sup>\*</sup>DIAGNOGENE

83, avenue Charles de Gaulle  
15000 Aurillac

<sup>\*\*</sup>LIMOS, UMR 6158 CNRS  
Univ. Clermont-Ferrand II  
63177 Aubière

**Résumé.** La notion de règles entre attributs est très générale, allant des règles d'association en fouille de données aux dépendances fonctionnelles (DF) en bases de données. Malgré cette diversité, la syntaxe des règles est toujours la même, seule leur sémantique diffère. Pour une sémantique donnée, en fonction des propriétés induites, des techniques algorithmiques sont mises en oeuvre pour découvrir les règles à partir des données. A partir d'un ensemble de règles, il est aussi utile en pratique de *raisonner* sur ces règles, comme cela est le cas par exemple avec les axiomes d'Armstrong pour les dépendances fonctionnelles. Dans cet article, nous proposons un cadre qui permet de s'assurer qu'une sémantique donnée pour les règles est *bien-formée*, i.e. les axiomes d'Armstrong sont justes et complets pour cette sémantique. Les propositions faites dans ce papier proviennent du contexte applicatif de l'*analyse de données de biopuces*. A partir de plusieurs sémantiques pour les données d'expression de gènes, nous montrons comment ces sémantiques s'intègrent dans le cadre présenté.

## 1 Introduction

Les biopuces permettent aujourd'hui aux biologistes de mesurer l'expression de milliers de gènes simultanément et un des défis majeurs fixé à présent est de comprendre les *réseaux de régulation géniques*, i.e. de découvrir les interactions entre les différents gènes.

Dans le cadre de nos travaux, nous nous intéressons à définir des *règles* entre les gènes à partir de données d'expression de gènes, étant entendu que ces règles forment un modèle possible d'un réseau de régulation. Ces données sont à valeurs réelles, chaque valeur représentant le niveau d'expression d'un gène pour une expérience (ou biopuce) particulière.

La notion de *règles* entre attributs est très générale, allant des règles d'association en fouille de données aux dépendances fonctionnelles en bases de données. Malgré cette diversité, la syntaxe des règles est toujours la même, seule leur sémantique diffère. Pour une sémantique donnée, en fonction des propriétés induites, des techniques algorithmiques sont mises en oeuvre pour découvrir les règles à partir des données [Agrawal et Srikant, 1994, Lopes *et al.*, 2002, Morishita et Sese, 2000].

A partir d'un ensemble de règles, il est aussi très utile en pratique de pouvoir