

Différentes sémantiques pour les règles : une définition aux propriétés intéressantes

Marie Agier

DIAGNOGENE SA, 15000 Aurillac, FRANCE
LIMOS, UMR 6158 CNRS, Univ. Clermont-Ferrand II, FRANCE

La notion de *règles* est très populaire en bases de données et en data mining. Plusieurs *sémantiques* pour les règles ont été définies jusque là, les deux plus connues étant les dépendances fonctionnelles et les règles d'association. Nous nous intéressons ici aux sémantiques pour lesquelles le système d'axiomes d'Armstrong est juste et complet, que nous appelons *sémantiques bien-formées*. Ces sémantiques vérifient des propriétés intéressantes, comme la possibilité de raisonner sur les règles ou de calculer des couvertures de l'ensemble des règles satisfaites.

Nous présentons ici une *définition générique* des sémantiques et nous montrons qu'une équivalence existe entre des restrictions syntaxiques sur cette définition et le fait que ces sémantiques soient bien-formées. D'un point de vue pratique, cette équivalence permet de prouver rapidement qu'une sémantique est bien-formée puisqu'il suffit de montrer qu'elle est en accord avec ces restrictions syntaxiques.

Nous montrons également la relation entre la définition des sémantiques proposée et le problème sous-jacent de découverte de connaissances, i.e. à partir d'une sémantique et d'une relation données, découvrir une couverture des règles satisfaites dans cette relation.

Ce travail prend ces racines dans l'analyse de données de biopuces, plus précisément la découverte de *réseaux de régulation de gènes*. Le challenge est de trouver les relations entre les gènes i.e. comment le niveau d'expression de chaque gène affecte le niveau d'expression des autres gènes. Nous pensons qu'il est important d'étudier plusieurs types de règles compte tenu des divers objectifs biologiques et qu'il n'est pas souhaitable de se restreindre à une seule sémantique (par exemple aux règles d'association). L'objectif est donc d'offrir aux biologistes un cadre dans lequel ils peuvent choisir ou définir les sémantiques qui correspondent au mieux à leurs attentes. Une fois que ces sémantiques sont prouvées bien-formées, i.e. elles sont simplement en accord avec les restrictions syntaxiques proposées, alors les méthodes d'inférence classiques sur les règles peuvent s'appliquer pour les biologistes.