

Causal Discovery in Non-Ideal Frameworks

Stijn Meganck^{*}, Philippe Leray⁺, Bernard Manderick^{*}

^{*} CoMo, Computational Modeling Lab
Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel
smeganck@vub.ac.be and <http://como.vub.ac.be>

⁺ LINA, Laboratoire d'Informatique de Nantes Atlantique
Polytech'Nantes, La Chantrerie, rue Christian Pauc, BP 50609, 44306
Nantes Cedex 3
philippe-leray@univ-nantes.fr et
<http://www.univ-nantes.fr>

Résumé

Les réseaux bayésiens causaux (RBC) sont un formalisme bien connu pour la représentation de relations causales dans un ensemble de variables. La découverte de ces relations causales passe généralement par l'utilisation de données d'expérimentation puisque l'observation seule de ces variables ne permet d'identifier que le graphe partiellement orienté, représentant de la classe d'équivalence de Markov. Cette découverte de relations causales à partir de données est un atout non négligeable dans de nombreux domaines d'application. La plupart des algorithmes réalisant cette tâche font deux hypothèses classiques qui facilitent grandement la découverte de causalité : tout d'abord toutes les variables du modèle sont déterminées, et l'on possède ensuite un nombre suffisant de mesures de ces variables. Dans nos travaux, nous partons des méthodes existantes dans ce cadre idéal comme point de départ pour de nouveaux algorithmes fonctionnant dans des situations moins idéales. Nous présentons ainsi deux algorithmes UnCaDo, qui peut être utilisé lorsque le nombre de données initiales est trop faible pour réaliser des tests statistiques fiables et MyCaDo++, qui peut être utilisé pour découvrir la présence de variables latentes. Ces algorithmes permettent alors de proposer des plans d'expériences permettant de déterminer la structure causale sous-jacente.

Mots-clés : Causalité, Apprentissage, Modèles Graphiques