

Fouille de règles différentielles causales dans les graphes de connaissances

Lucas Simonne* , Nathalie Pernelle** , Fatiha Saïs*

*LRI, Université Paris Saclay
simonne@lri.fr, saïs@lri.fr

**LIPN, Université Sorbonne Paris Nord
pernelle@lipn.univ-paris13.fr

Résumé. L'intérêt porté aux graphes de connaissances a grandi ces dernières années, à la fois dans le milieu académique et industriel. Nous proposons une approche permettant de déterminer des règles différentielles causales dans des graphes de connaissances. Ces règles expriment qu'un traitement différent pour deux instances d'une classe conduit à un résultat différent pour ces instances. L'approche que nous proposons est basée sur des appariements sémantiques, et des strates définies comme étant des sous-classes complexes. Une première expérimentation sur un extrait de DBPedia montre que de telles règles peuvent permettre d'expliquer certaines variabilités dans les propriétés étudiées.

1 Introduction

Les graphes de connaissances permettent de stocker des données et des connaissances sur des différentes thématiques, ie. expériences scientifiques, politique, finance, etc. Ce type de données est aujourd'hui utilisé dans le milieu académique ou industriel. De nombreux travaux de recherche récents ont permis d'effectuer de la fouille de règles plus ou moins expressives dans les graphes de connaissances (Galárraga et al. (2013)). Ces règles permettent de corriger des faits erronés ou encore de déduire des faits manquants. Cependant, il n'existe que très peu d'approches qui ont pour objectif de découvrir des règles causales dans de tels graphes.

Un grand nombre d'expérimentations sont menées afin d'étudier l'existence de liens causaux, et ce quel que soit le domaine, qu'il s'agisse de médecine, d'environnement ou encore de marketing. De nombreux travaux de recherche ont pour objectif de découvrir des règles causales dans des données relationnelles. Ces travaux sont basés sur les modèles causaux structurels (Pearl (2009)), les réseaux bayésiens (Neapolitan (2003)) ou encore sur les modèles à résultats potentiels - i.e. *potential outcome* en anglais (Rubin D. B (1974)). Une approche standard consiste à construire des expériences où l'attribution d'un traitement est aléatoire afin de déterminer son effet au sein d'une population. Ces expériences sont toutefois compliquées à mettre en place, pour des raisons multiples et en particulier le coût de ces expériences ou encore les problèmes éthiques qu'elles