

Enrichissement de règles de Horn par des prédicats numériques

Armita Khajeh Nassiri*, Nathalie Pernelle**, Fatiha Saïs*

* LISN, CNRS (UMR 9015), Université Paris Saclay , France

** LIPN, CNRS (UMR 7030), Université Sorbonne Paris Nord, France
firstname.lastname@lri.fr

Résumé. Dans cet article nous présentons REGNUM, un système qui enrichit le corps de règles déjà découvertes dans un graphe de connaissances avec des atomes impliquant des prédicats numériques dont les valeurs sont contraintes par des intervalles spécifiés. Les intervalles sont obtenus à l'aide de techniques de discrétisation supervisées, avec l'objectif d'augmenter la confiance des règles fournies par la méthode de découverte de règles. Nos résultats expérimentaux démontrent que les règles enrichies avec des prédicats numériques sont de meilleure qualité globale et sont mieux adaptées pour la tâche d'enrichissement de graphes de connaissances comparativement aux règles initiales.

1 Introduction

Les graphes de connaissances (GC) représentent des faits sur de nombreuses entités du monde exprimés dans un format interprétable par les machines. Ces graphes intègrent différentes formes de connaissances, et de nombreux travaux ont été consacrés à l'acquisition de ces connaissances. Un type d'approches sont celles permettant la fouille de règles logiques dans les GCs. Ces règles peuvent servir à compléter le GC, à détecter des données erronées et à aligner des ontologies.

AMIE(Lajus et al., 2020) est un système de fouille de règles pour les GCs très connu dans l'état de l'art. Il est efficace et exhaustif, c'est-à-dire qu'il extrait toutes les règles connectées et fermées en fonction de seuils définis sur des mesures de qualité (par exemple, la confiance et la couverture des têtes des règles) et d'un nombre maximal d'atomes spécifié. AMIE peut découvrir des règles qui impliquent des constantes (par exemple, $age(x, 53)$). Cependant, ces règles peuvent être trop spécifiques et peu intéressantes lorsqu'il s'agit de ces prédicats. RuDiK (Ortona et al., 2018) propose une approche non exhaustive pour découvrir des règles logiques plus expressives. RuDiK peut prédire l'absence d'un fait et permet d'effectuer des comparaisons au-delà des égalités en utilisant les relations appartenant à l'ensemble $rel \in \{<, \leq, \neq, \geq, >\}$. Les valeurs impliquées ne sont pas des seuils mais proviennent du graphe de connaissances lui-même.

Il existe également des systèmes de fouille de règles tels que AnyBURL (Meilicke et al., 2019) se focalise uniquement sur les règles fondées sur des chemins dans le graphe. AnyBURL est une approche ascendante qui commence par l'échantillonnage de chemins spécifiques dans