Mach Learn
DOI 10.1007/s10994-013-5402-3

Finding relational redescriptions

Esther Galbrun · Angelika Kimmig

Received: 13 December 2012 / Accepted: 25 July 2013

Abstract We introduce relational redescription mining, that is, the task of finding two structurally different patterns that describe nearly the same set of object pairs in a relational dataset. By extending redescription mining beyond propositional and real-valued attributes, it provides a powerful tool to match different relational descriptions of the same concept.

We propose an alternating scheme for solving this problem. Its core consists of a novel relational query miner that efficiently identifies discriminative connection patterns between pairs of objects. Compared to a baseline Inductive Logic Programming (ILP) approach, our query miner is able to mine more complex queries, much faster. We performed extensive experiments on three real world relational datasets, and present examples of redescriptions found, exhibiting the power of the method to expressively capture relations present in these networks.

Keywords Redescription mining · Relational query mining · Inductive Logic Programming · Graph mining · Relational data mining

1 Introduction

With the increasing amount of data available from heterogenous sources nowadays, establishing links between different perspectives on the same concept becomes ever more important, as recognized, for instance, in schema matching and ontology alignment for the

Editors: Fabrizio Riguzzi and Filip Železný.

E. Galbrun (✉)
Department of Computer Science and Helsinki Institute for Information Technology HIIT, University of Helsinki, P.O. Box 68, 00014 Helsinki, Finland
e-mail: esther.galbrun@cs.helsinki.fi

A. Kimmig
Departement Computerwetenschappen, KU Leuven, Celestijnenlaan 200A, bus 2402, 3001 Heverlee, Belgium
e-mail: angelika.kimmig@cs.kuleuven.be

Published online: 31 August 2013