

Learning graphical models for relational data via lattice search

Oliver Schulte · Hassan Khosravi

Received: 28 November 2010 / Accepted: 26 March 2012 / Published online: 30 May 2012
© The Author(s) 2012

Abstract Many machine learning applications that involve relational databases incorporate first-order logic and probability. Relational extensions of graphical models include Parametrized Bayes Net (Poole in IJCAI, pp. 985–991, 2003), Probabilistic Relational Models (Getoor et al. in Introduction to statistical relational learning, pp. 129–173, 2007), and Markov Logic Networks (MLNs) (Domingos and Richardson in Introduction to statistical relational learning, 2007). Many of the current state-of-the-art algorithms for learning MLNs have focused on relatively small datasets with few descriptive attributes, where predicates are mostly binary and the main task is usually prediction of links between entities. This paper addresses what is in a sense a complementary problem: learning the structure of a graphical model that models the distribution of discrete descriptive attributes given the links between entities in a relational database. Descriptive attributes are usually nonbinary and can be very informative, but they increase the search space of possible candidate clauses. We present an efficient new algorithm for learning a Parametrized Bayes Net that performs a level-wise search through the table join lattice for relational dependencies. From the Bayes net we obtain an MLN structure via a standard moralization procedure for converting directed models to undirected models. Learning MLN structure by moralization is 200–1000 times faster and scores substantially higher in predictive accuracy than benchmark MLN algorithms on five relational databases.

Keywords Statistical-relational learning · Graphical models · Markov logic networks · Bayes nets

1 Introduction

Many databases store data in relational format, with different types of entities and information about links between the entities. The field of statistical-relational learning (SRL)

Editor: Lise Getoor.

O. Schulte (✉) · H. Khosravi
School of Computing Science, Simon Fraser University, Vancouver-Burnaby, BC, V5A 1S6, Canada
e-mail: oschulte@cs.sfu.ca

H. Khosravi
e-mail: hkhosrav@cs.sfu.ca