Mach Learn (2014) 94:105–125
DOI 10.1007/s10994-013-5362-7

# Modelling relational statistics with Bayes Nets

**Oliver Schulte · Hassan Khosravi · Arthur E. Kirkpatrick · Tianxiang Gao · Yuke Zhu**

Received: 11 December 2012 / Accepted: 13 April 2013 / Published online: 1 May 2013


**Abstract** Class-level models capture relational statistics over object attributes and their connecting links, answering questions such as "what is the percentage of friendship pairs where both friends are women?" Class-level relationships are important in themselves, and they support applications like policy making, strategic planning, and query optimization. We represent class statistics using Parametrized Bayes Nets (PBNs), a first-order logic extension of Bayes nets. Queries about classes require a new semantics for PBNs, as the standard grounding semantics is only appropriate for answering queries about specific ground facts. We propose a novel random selection semantics for PBNs, which does not make reference to a ground model, and supports class-level queries. The parameters for this semantics can be learned using the recent pseudo-likelihood measure (Schulte in SIAM SDM, pp. 462–473, 2011) as the objective function. This objective function is maximized by taking the empirical frequencies in the relational data as the parameter settings. We render the computation of these empirical frequencies tractable in the presence of negated relations by the inverse Möbius transform. Evaluation of our method on four benchmark datasets shows that maximum pseudo-likelihood provides fast and accurate estimates at different sample sizes.

**Keywords** Statistical-relational learning · Bayes Nets · Structured data · Pseudo-likelihood · Möbius transform

## 1 Introduction

Many applications store data in relational form. Relational data introduces the machine learning problem of *class-level frequency estimation*: building a model that can answer statistical queries about classes of individuals in the database (Getoor et al. 2001). Examples of such queries include:

Editors: Fabrizio Riguzzi and Filip Zelezny.

O. Schulte (✉) · H. Khosravi · A.E. Kirkpatrick · T. Gao · Y. Zhu
School of Computing Science, Simon Fraser University, Burnaby, British Columbia, Canada
e-mail: oschulte@cs.sfu.ca