

Learning Probabilistic Models from Relational Data

—Invited Lecture—

Daphne Koller
Stanford University

Abstract

Bayesian networks are a compact and natural representation for complex probabilistic models. They use graphical notation to encode domain structure: the direct probabilistic dependencies between variables in the domain. However, many real-world domains are best described by relational domains in which instances of multiple types are related to each other in complex ways. For example, in a scientific paper domain, papers are related to each other via citation, and are also related to their authors. Bayesian networks are attribute-based, making it difficult to represent the rich relational structure of complex domains involving multiple entities that interact with each other. The talk will describe probabilistic relational models (PRMs), a new probabilistic modeling language suitable for relational domains. PRMs extend the language of Bayesian networks with the expressive power of object-relational languages. They model the uncertainty over the attributes of objects in the domain as well as uncertainty over the relations between objects. The talk will present techniques for automatically inducing PRMs directly from a relational data set, and applications of these techniques to pattern discovery in complex real-world data sets, and to clustering and classification in relational domains.

Joint work with Lise Getoor, Nir Friedman, Avi Pfeffer, Eran Segal, and Ben Taskar