Learning the structure of a Bayesian network: An application of Information Geometry and the Minimum Description Length Principle

Eitel Lauria, PhD
School of Computer Science and Mathematics, Marist College
Poughkeepsie, NY, USA
Eitel.Lauria@marist.edu
http://academic.marist.edu/~jf4n

Abstract

The field of Bayesian Networks has had an enormous development over the last few years and is one of the current key topics of research in the design of statistical machine learning and data mining algorithms. Bayesian networks are a natural marriage between two areas in mathematics: graph theory and probability theory. A Bayesian net encodes the probability distribution of a set of attributes by specifying a set of conditional independence assumptions together with a set of relationships among these attributes and their related joint probabilities. When used in this way, Bayesian networks result in a powerful knowledge representation formalism based on probability and provide a natural way of dealing with uncertainty and complexity, two recurring topics that have impact across a wide range of knowledge domains. The present paper addresses the issue of learning the underlying model of the Bayesian network, expressed as a digraph, which includes the specification of the conditional independence assumptions among the attributes of the model; and given the model, the conditional probability distributions that quantify those dependencies. We heuristically search the space of network structures using a scoring function based on the updated version of the Minimum Description Length Principle, that takes into account the volume of the model manifold [1] [2]. We present empirical results on synthetic datasets that analyse the relative effectiveness of this approach when varying the size and complexity of a Bayesian network.

References:

[1] J. Rissanen (1996), "Fisher Information and Stochastic Complexity", IEEE Transaction on Information Theory, 42, 40-47

[2] C. Rodriguez (2001), "Entropic priors for discrete probabilistic networks and for mixtures of Gaussian models", presented at MaxEnt2001, APL Johns Hopkins University, August 4-9 2001