Structure Inference of Bayesian Networks from Data: A New Approach Based on Generalized Conditional Entropy

Dan A. Simovici*, Saaid Baraty*

*Univ. of Massachusetts Boston, Massachusetts 02125, USA
{dsim,sbaraty}@cs.umb.edu

Abstract. We propose a novel algorithm for extracting the structure of a Bayesian network from a dataset. Our approach is based on generalized conditional entropies, a parametric family of entropies that extends the usual Shannon conditional entropy. Our results indicate that with an appropriate choice of a generalized conditional entropy we obtain Bayesian networks that have superior scores compared to similar structures obtained by classical inference methods.

1 Introduction

A Bayesian Belief Network (BBN) structure is a directed acyclic graph which represents probabilistic dependencies among a set of random variables.

Inducing a BBN structure for the set of attributes of a dataset is a well known problem and a challenging one due to enormity of the search space. The number of possible BBN structures grows super-exponentially with respect to the number of the nodes.

In Cooper and Herskovits (1993), where the K2 heuristic algorithm is introduced, a measure of the quality of the structure is derived based on its posterior probability in presence of a dataset. An alternative approach to compute a BBN structure is based on the Minimum Description Length principle (MDL) first introduced in Rissanen (1978). The algorithms of Lam and Bacchus (1994) and Suzuki (1999) are derived from this principle.

We propose a new approach to inducing BBN structures from datasets based on the notion of β-generalized entropy (β-GE) and its corresponding β-generalized conditional entropy (β-GCE) introduced in Havrda and Charvat (1967) and axiomatized in Simovici and Jaroszewicz (2002) as a one-parameter family of functions defined on partitions (or probability distributions). The flexibility that ensues allows us to generate BBNs with better scores than published results.

One important advantage of our approach is that, unlike Cooper and Herskovits (1993) it is not based on any distributional assumption for developing the formula.

2 Generalized Entropy and Structure Inference

The set of partitions of a set S is denoted by $\mathsf{PART}(S)$. The *trace of a partition* π on a subset T of S is the partition $\pi_T = \{T \cap B_i \mid i \in I \text{ and } T \cap B_i \neq \emptyset\}$ of T. The usual order between set partitions is denoted by "$\leq$". It is well-known that $(\mathsf{PART}(S), \leq)$ is a bounded