Evaluating Bayesian Networks by Sampling with Simplified Assumptions

Saaid Baraty*, Dan A. Simovici*

*University of Massachusetts Boston
Computer Science Department,
Boston, Massachusetts 02125
e-mail{sbaraty,dsim}@cs.umb.edu,

Abstract. The most common fitness evaluation for Bayesian networks in the presence of data is the Cooper-Herskovitz criterion. This technique involves massive amounts of data and, therefore, expansive computations. We propose a cheaper alternative evaluation method using simplified assumptions which produces evaluations that are strongly correlated with the Cooper-Herskovitz criterion.

1 Introduction

We investigate the problem of constructing a Bayesian network for a composite phenomenon $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ where $\mathbf{u}_i$ for $1 \leq i \leq n$ are discrete random variables representing the state assignment of the attributes of $\mathbf{U}$. To accomplish this, we start from a data multiset $\mathcal{D} = \{t_1, t_2, \ldots, t_m\}$ where an n-ary tuple t_i is an instance of the event $\mathbf{U}$. We refer to this multiset as *evidence data set* (data set for short).

A number of assumptions are necessary for deriving a measure for evaluating the fitness of a Bayesian network structure (BNS) for a training data set. Stronger hypotheses make the evaluation more manageable. On the other hand, the model obtained under weaker assumptions is better capable to be conforming with the underlying true distribution of the problem.

Let $G = (\mathbf{U}, E)$ be a directed acyclic graph having $\mathbf{U}$ as its set of vertices and E as its set of edges, which captures the direct probabilistic dependencies among these variables. Let Θ be the collection of parameters which quantifies the joint probability distribution of $\mathbf{U}$ as specified by G. We denote the set of possible assignments of a random variable $\mathbf{u}_i$ by $\mathrm{Dom}(\mathbf{u}_i) = \{u_i^1, \ldots, u_i^{r_i}\}$. The notion of domain can be extended to sets of variables $\mathbf{V}$ using Cartesian product. If the *set of parent nodes* of $\mathbf{u}_i$ is $\mathsf{Par}_G(\mathbf{u}_i)$, then $\mathrm{Dom}(\mathsf{Par}_G(\mathbf{u}_i) = \{U_i^1, \ldots, U_i^{q_i}\}$. The set of *non-descendants of* $\mathbf{u}_i$, $\mathsf{nd}_G(\mathbf{u}_i)$ is the set of all nodes in $\mathbf{U}$ excluding $\mathbf{u}_i$ and all its descendants. When it is clear from the context we drop the subscript G. The pair $\mathcal{B} = (G, \Theta)$ satisfies the *local Markov condition* if $P_{\mathcal{B}}(\mathbf{u}_i|\mathsf{nd}(\mathbf{u}_i)) = P_{\mathcal{B}}(\mathbf{u}_i|\mathsf{Par}(\mathbf{u}_i))$ for $1 \leq i \leq n$, where $P_{\mathcal{B}}$ is the probability distribution on $\mathbf{U}$ specified by $\mathcal{B}$. The model $\mathcal{B}$ is a *Bayesian* network if it satisfies the local Markov condition. By the chain rule we have: $P_{\mathcal{B}}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \prod_{i=1}^{n} P_{\mathcal{B}}(\mathbf{u}_i|\mathsf{Par}(\mathbf{u}_i))$. Therefore if we let $\theta_{ijk} = P(\mathbf{u}_i = u_i^k|\mathsf{Par}(\mathbf{u}_i) = U_i^j)$ and $\theta_{ij\cdot} = (\theta_{ij1}, \ldots, \theta_{ijr_i})$ for $1 \leq i \leq n$, $1 \leq k \leq r_i$ and $1 \leq j \leq q_i$, then the joint probability distribution on $\mathbf{U}$ is specified by $\Theta = \{\theta_{ij\cdot}|1 \leq i \leq n \text{ and } 1 \leq j \leq q_i\}$.