

# An optimization-based approach for the design of Bayesian networks

Ana M. Martínez-Rodríguez<sup>a</sup>, Jerrold H. May<sup>b</sup>, Luis G. Vargas<sup>b,\*</sup>

<sup>a</sup> University of Jaen, Spain

<sup>b</sup> The Joseph M Katz Graduate School of Business, University of Pittsburgh, United States

Received 26 November 2007; accepted 2 January 2008

---

## Abstract

Bayesian networks model conditional dependencies among the domain variables, and provide a way to deduce their interrelationships as well as a method for the classification of new instances. One of the most challenging problems in using Bayesian networks, in the absence of a domain expert who can dictate the model, is inducing the structure of the network from a large, multivariate data set. We propose a new methodology for the design of the structure of a Bayesian network based on concepts of graph theory and nonlinear integer optimization techniques.

© 2008 Elsevier Ltd. All rights reserved.

*Keywords:* Bayesian networks; Optimal design

---

## 1. Introduction

A Bayesian network is a graphical representation of an  $n$ -dimensional probability distribution. It is a directed acyclic graph (DAG) in which each node represents a variable of interest, and the arcs represent dependencies between the variables. The strengths of the dependencies are quantified by conditional probabilities. Thus, the architecture of a Bayesian network involves two components: (1) the structure of the network, which describes the direct association pairs of variables, and (2) the parameters of the network that represent the probability distribution. In some contexts, the structure of the network may have a causal interpretation, i.e., the ending node of an arc may be a direct effect of the node at the beginning of the arc. In such cases, a Bayesian network is known as a *causal network* [33]. To represent causal relations, we identify the parents of each node (nodes incident to the tails of all arcs into the node) that are the direct causes of the node. Bayesian networks have been used in a wide variety of domains, including biology [18], computer games [42], data mining [29,4], diagnosis of failures [5,37], medicine [38,3,2,28,43], prediction [1,44], reliability analysis [41], students advising [13,35] and weather forecasting [24].

There are several advantages in representing a probability distribution by a Bayesian network. Bayesian networks allow a fast and intuitive understanding of the relations (dependence vs. independence) between the variables. The chain-rule allows joint distributions to be represented as a product of conditional distributions. In order to use

---

\* Corresponding author.

E-mail address: [lgvargas@pitt.edu](mailto:lgvargas@pitt.edu) (L.G. Vargas).

a Bayesian network, it is necessary to have at least a subset of the conditional distributions available. However, in general, it is less difficult to derive those conditional distributions than it is to specify the joint distribution. The probability distribution represented by a Bayesian network can be used, for example, to do Bayesian inference, i.e., if values of some of the variables are known and they are introduced in the network, it is possible to use the network to compute other probabilities, given partial evidence.

Two approaches are available for determining the structure of a Bayesian network. One, a human expert who is an expert in the domain dictates the arcs that connect the nodes in the network. Two, a computerized induction method is used to derive a structure directly from a relevant data set. Construction of Bayesian networks using the former method is time consuming and both benefits from, and is limited by, the subjective judgment of the expert consulted. For those reasons, there has been substantial recent interest in identifying a computerized, objective methodology for inducing the graphical structure of the Bayesian network from a multivariate data set. We propose a new approach, using a mathematical program, to create the network structure. In order to formulate the objective function of that mathematical program, we define coefficients of dependence between variables. The coefficients are used to define a measure of the global dependence in the network, which is then maximized over the space of structures of Bayesian networks.

## 2. Background

A large number of algorithms have been proposed to learn Bayesian network structures. They can be classified according to the nature of the modeling into Score + Search methods and Detection of Conditional (In)Dependencies.

1. Score + Search methods. These methods use a metric to measure the goodness-of-fit of every candidate Bayesian network with respect to a database of cases, and a search procedure to move through the space of possible network structures. According to the score metrics, the models are classified into:
  - (a) Bayesian Scoring: In this case a Bayesian approach is considered. Thus, given a database over a set of  $n$  variables, the model that maximizes the posterior probability is selected [15,21,25,30,16,9,20].
  - (b) Information theory Scoring: In this case the scores are based on the entropy of the network. The Kullback–Leibler cross-entropy measure is the score most frequently used [12,23,36].
2. Detection of Conditional (In)Dependencies method. This approach is known as constraint-based learning. The algorithms attempt to recover the structure of the Bayesian network by detecting the conditional (in)dependencies among the variables in the data [19,8,6].

The search for the best network with respect to some criterion is performed in the space of all possible networks, and the number of elements in this space increases exponentially with the number of nodes (or variables) [34], finding the best structure is NP-hard [10,11]. Thus, the use of search heuristics that look for a network structure that is good is justified [31,39]. However, it has not been demonstrated that the solution found by these algorithms is the optimal one, and/or the degree of closeness of the generated solution to the optimal one, i.e., the heuristic worst-case performance ratio.

On the other hand, conditional (in)dependency is a concept to deal with probability distributions,  $P(\cdot)$  that admit a representation by a graph  $G$  that is a perfect map of  $P(\cdot)$ , i.e.,  $G$  entails only conditional independencies in  $P$  and all conditional independencies in  $P$  are entailed by  $G$ , based on the Markov condition. In general, the proposed algorithms following this methodology have as input some conditional (in)dependencies that can be obtained from the data, and not a list of all the conditional (in)dependencies that admit a representation by means of a perfect map, because some of the structures obtained may not be directed acyclic graphs. In this paper we depart from both methodologies in an attempt to correct some of their disadvantages.

## 3. Coefficients of dependence

Consider a finite set of variables  $V = (V_1, \dots, V_n)$ , one or more of which is distinguished as being dependent, to be predicted by the remaining independent variables. For example, in marketing, the dependent variables might be demographic characteristics of a television viewer such as age, family income, and educational level, and the independent variables might be whether or not the person watched each of the 240 most popular broadcast programs over the previous week. In such a domain, the purpose of the Bayesian network is to build a model that could be used

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات