Chapter 1.16
Learning Bayesian Networks

Marco F. Ramoni
Harvard Medical School, USA

Paola Sebastiani
Boston University School of Public Health, USA

INTRODUCTION

Born at the intersection of artificial intelligence, statistics, and probability, Bayesian networks (Pearl, 1988) are a representation formalism at the cutting edge of knowledge discovery and data mining (Heckerman, 1997). Bayesian networks belong to a more general class of models called *probabilistic graphical models* (Whittaker, 1990; Lauritzen, 1996) that arise from the combination of graph theory and probability theory, and their success rests on their ability to handle complex probabilistic models by decomposing them into smaller, amenable components. A probabilistic graphical model is defined by a graph, where nodes represent stochastic variables and arcs represent dependencies among such variables. These arcs are annotated by probability distribution shaping the interaction between the linked variables. A probabilistic graphical model is called a Bayesian network, when the graph connecting its variables is a directed acyclic graph (DAG). This graph represents conditional independence assumptions that are used to factorize the joint probability distribution of the network variables, thus making the process of learning from a large database amenable to computations. A Bayesian network induced from data can be used to investigate distant relationships between variables, as well as making prediction and explanation, by computing the conditional probability distribution of one variable, given the values of some others.

BACKGROUND

The origins of Bayesian networks can be traced back as far as the early decades of the 20th century, when Sewell Wright developed path analysis to aid the study of genetic inheritance (Wright, 1923, 1934). In their current form, Bayesian networks were introduced in the early 1980s as a knowledge representation formalism to encode and use the information acquired from human experts in automated reasoning systems in order to perform diagnostic, predictive, and explanatory tasks

(Charniak, 1991; Pearl, 1986, 1988). Their intuitive graphical nature and their principled probabilistic foundations were very attractive features to acquire and represent information burdened by uncertainty. The development of amenable algorithms to propagate probabilistic information through the graph (Lauritzen, 1988; Pearl, 1988) put Bayesian networks at the forefront of artificial intelligence research. Around the same time, the machine-learning community came to the realization that the sound probabilistic nature of Bayesian networks provided straightforward ways to learn them from data. As Bayesian networks encode assumptions of conditional independence, the first machine-learning approaches to Bayesian networks consisted of searching for conditional independence structures in the data and encoding them as a Bayesian network (Glymour, 1987; Pearl, 1988). Shortly thereafter, Cooper and Herskovitz (1992) introduced a Bayesian method that was further refined by Heckerman, et al. (1995) to learn Bayesian networks from data.

These results spurred the interest of the data-mining and knowledge-discovery community in the unique features of Bayesian networks (Heckerman, 1997); that is, a highly symbolic formalism, originally developed to be used and understood by humans, well-grounded on the sound foundations of statistics and probability theory, able to capture complex interaction mechanisms and to perform prediction and classification.

MAIN THRUST

A Bayesian network is a graph, where nodes represent stochastic variables and (arrowhead) arcs represent dependencies among these variables. In the simplest case, variables are discrete, and each variable can take a finite set of values.

Representation

Suppose we want to represent the variable *gender*. The variable gender may take two possible values: male and female. The assignment of a value to a variable is called the *state of the variable*. So, the variable gender has two states: Gender = Male and Gender = Female. The graphical structure of a Bayesian network looks like this:

The network represents the notion that obesity and gender affect the heart condition of a patient. The variable obesity can take three values: yes, borderline and no. The variable heart condition has two states: true and false. In this representation, the node heart condition is said to be a *child* of the nodes gender and obesity, which, in turn, are the *parents* of heart condition.

Figure 1.