B

Bayesian Networks

Ahmad Bashir
University of Texas at Dallas, USA

Latifur Khan
University of Texas at Dallas, USA

Mamoun Awad
University of Texas at Dallas, USA

INTRODUCTION

A *Bayesian network* is a graphical model that finds probabilistic relationships among variables of a system. The basic components of a Bayesian network include a set of nodes, each representing a unique variable in the system, their inter-relations, as indicated graphically by edges, and associated probability values. By using these probabilities, termed *conditional probabilities,* and their interrelations, we can reason and calculate unknown probabilities. Furthermore, Bayesian networks have distinct advantages compared to other methods, such as neural networks, decision trees, and rule bases, which we shall discuss in this paper.

BACKGROUND

Bayesian classification is based on Naïve Bayesian classifiers, which we discuss in this section. *Naive Bayesian classification* is the popular name for a probabilistic classification. The term Naive Bayes refers to the fact that the probability model can be derived by using Bayes' theorem and that it incorporates strong independence assumptions that often have no bearing in reality, hence they are deliberately naïve. Depending on the model, Naïve Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for Naïve Bayes models uses the method of maximum likelihood.

Abstractly, the desired probability model for a classifier is a conditional model

$$P(C \mid F_1, \ldots, F_n)$$

over a dependent class variable C with a small number of outcomes, or classes, conditional on several feature variables F_1 through F_n. The naïve conditional independence assumptions play a role at this stage, when the probabilities are being computed. In this model, the assumption is that each feature F_i is conditionally independent of every other feature F_j. This situation is mathematically represented as:

$$P(F_i \mid C, F_j) = P(F_i \mid C)$$

Such naïve models are easier to compute, because they factor into P(C) and a series of independent probability distributions. The Naïve Bayes classifier combines this model with a decision rule. The common rule is to choose the label that is most probable, known as the maximum a posteriori or MAP decision rule.

In a supervised learning setting, one wants to estimate the parameters of the probability model. Because of the independent feature assumption, it suffices to estimate the class prior and the conditional feature models independently by using the method of maximum likelihood.

The Naïve Bayes classifier has several properties that make it simple and practical, although the independence assumptions are often violated. The overall classifier is the robust to serious deficiencies of its underlying naïve probability model, and in general, the Naïve Bayes approach is more powerful than might be expected from the extreme simplicity of its model; however, in the presence of nonindependent attributes w_i, the Naïve Bayesian classifier must be upgraded to the Bayesian classifier, which will more appropriately model the situation.

MAIN THRUST

The basic concept in the Bayesian treatment of certainties in causal networks is *conditional probability*. When the probability of an event *A*, *P*(*A*), is known, then it is conditioned by other known factors. A conditional probability statement has the following form:

Given that event B has occurred, the probability of the event A occurring is x.

Graphical Models

A graphical model visually illustrates conditional independencies among variables in a given problem. Two variables that are conditionally independent have no direct impact on each other's values. Furthermore, the graphical model shows any intermediary variables that separate two conditionally independent variables. Through these intermediary variables, two conditionally independent variables affect one another.

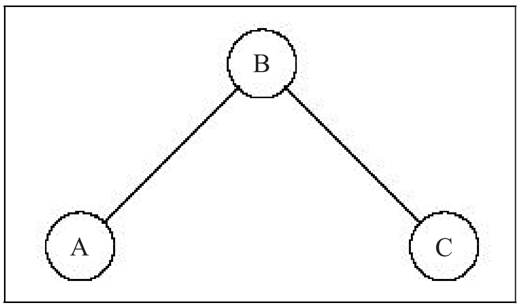
A graph is composed of a set of nodes, which represent variables, and a set of edges. Each edge connects two nodes, and an edge can have an optional direction assigned to it. For X_1 and X_2, if a causal relationship between the variables exists, the edge will be directional, leading from the case variable to the effect variable; if just a correlation between the variables exists, the edge will be undirected.

We use an example with three variables to illustrate these concepts. In this example, two conditionally independent variables, *A* and *C,* are directly related to another variable, *B.* To represent this situation, an edge must exist between the nodes of the variables that are directly related, that is, between *A* and *B* and between *B* and *C.* Furthermore, the relationships between *A* and *B* and *B* and *C* are correlations as opposed to causal relations; hence, the respective edges will be undirected. Figure 1 illustrates this example. Due to conditional independence, nodes *A* and *C* still have an indirect influence on one another; however, variable *B* encodes the information from *A* that impacts *C,* and vice versa.

A Bayesian network is a specific type of graphical model, with directed edges and no cycles (Stephenson, 2000). The edges in Bayesian networks are viewed as causal connections, where each parent node causes an effect on its children.

In addition, nodes in a Bayesian network contain a *conditional probability table,* or *CPT,* which stores all probabilities that may be used to reason or make inferences within the system.

Figure 1. Graphical model of two independent variables A and C that are directly related to a third variable B



Bayesian Probabilities

Probability calculus does not require that the probabilities be based on theoretical results or frequencies of repeated experiments, commonly known as *relative frequencies.* Probabilities may also be completely subjective estimates of the certainty of an event.

Consider an example of a basketball game. If one were to bet on an upcoming game between Team A and Team B, it is important to know the probability of Team A winning the game. This probability is definitely not a ratio, a relative frequency, or even an estimate of a relative frequency; the game cannot be repeated many times under exactly the same conditions. Rather, the probability represents only one's belief concerning Team A's chances of winning. Such a probability is termed a *Bayesian* or *subjective probability* and makes use of Bayes' theorem to calculate unknown probabilities.

A Bayesian probability may also be referred to as a personal probability. The Bayesian probability of an event *x* is a person's *degree of belief* in that event. A Bayesian probability is a property of the person who assigns the probability, whereas a *classical probability* is a physical property of the world, meaning it is the physical probability of an event.

An important difference between physical probability and Bayesian probability is that repeated trials are not necessary to measure the Bayesian probability. The Bayesian method can assign a probability for events that may be difficult to experimentally determine. An oft-voiced criticism of the Bayesian approach is that probabilities seem arbitrary, but this is a probability assessment issue that does not take away from the many possibilities that Bayesian probabilities provide.

Causal Influence

Bayesian networks require an operational method for identifying causal relationships in order for accurate domain modeling. Hence, causal influence is defined in the following manner: If the action of making variable *X* take some value sometimes changes the value taken by variable *Y,* then *X* is assumed to be responsible for sometimes changing *Y*'s value, and one may conclude that *X* is a cause of *Y.* More formally, *X* is manipulated when we force *X* to take some value, and we say *X* causes *Y* if some manipulation of *X* leads to a change in the probability distribution of *Y.*

Furthermore, if manipulating *X* leads to a change in the probability distribution of *Y,* then *X* obtaining a value by any means whatsoever also leads to a change in the probability distribution of *Y.* Hence, one can make the natural conclusion that causes and their effects are statis-

Related Content