P

Pseudo Independent Models

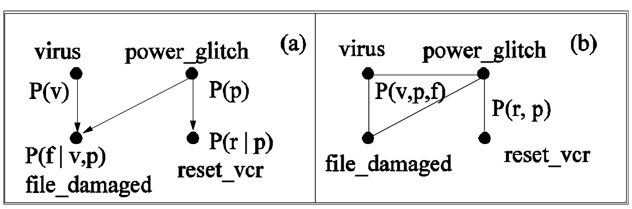
Yang Xiang
University of Guelph, Canada

INTRODUCTION

Graphical models such as Bayesian networks (BNs) (Pearl, 1988) and decomposable Markov networks (DMNs) (Xiang, Wong & Cercone, 1997) have been applied widely to probabilistic reasoning in intelligent systems. Figure1 illustrates a BN and a DMN on a trivial uncertain domain: A virus can damage computer files, and so can a power glitch. A power glitch also causes a VCR to reset. The BN in (a) has four nodes, corresponding to four binary variables taking values from {true, false}. The graph structure encodes a set of dependence and independence assumptions (e.g., that f is directly dependent on v, and p but is independent of r, once the value of p is known). Each node is associated with a conditional probability distribution conditioned on its parent nodes (e.g., $P(f \mid v, p)$). The joint probability distribution is the product $P(v, p, f, r) = P(f \mid v, p)\, P(r \mid p)\, P(v)\, P(p)$. The DMN in (b) has two groups of nodes that are maximally pair-wise connected, called *cliques*. Each clique is associated with a probability distribution (e.g., clique $\{v, p, f\}$ is assigned $P(v, p, f)$). The joint probability distribution is $P(v, p, f, r) = P(v, p, f)\, P(r, p) / P(p)$, where $P(p)$ can be derived from one of the clique distributions. The networks, for instance, can be used to reason about whether there are viruses in the computer system, after observations on f and r are made.

Construction of such networks by elicitation from domain experts can be very time-consuming. Automatic discovery (Neapolitan, 2004) by exhaustively testing all possible network structures is intractable. Hence, heuristic search must be used. This article examines a class of graphical models that cannot be discovered using the common heuristics.

Figure 1. (a) a trivial example BN; (b) a corresponding DMN



BACKGROUND

Let V be a set of n discrete variables $x_1, \ldots, x_n$ (in what follows, we will focus on finite, discrete variables). Each variable x_i has a finite space $S_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,D_i}\}$ of cardinality D_i. When there is no confusion, we write $x_{i,j}$ as x_{ij} for simplicity. The space of a set V of variables is defined by the Cartesian product of the spaces of all variables in V, that is, $S_V = S_1 \times \ldots \times S_n$ (or $\Pi_i S_i$). Thus, S_V contains the tuples made of all possible combinations of values of the variables in V. Each tuple is called a configuration of V, denoted by $\boldsymbol{v} = (x_1, \ldots, x_n)$.

Let $P(x_i)$ denote the probability function over x_i and $P(x_{ij})$ denote the probability value $P(x_i = x_{ij})$. A probabilistic domain model (PDM) M over V defines the probability values of every configuration for every subset $A \subseteq V$. Let $P(V)$ or $P(x_1, \ldots, x_n)$ denote the joint probability distribution (JPD) function over $x_1, \ldots, x_n$ and $P(x_{1j1}, \ldots, x_{njn})$ denote the probability value of a configuration $(x_{1j1}, \ldots, x_{njn})$. We refer to the function $P(A)$ over $A \subset V$ as the *marginal distribution* over A and $P(x_i)$ as the *marginal distribution* of x_i. We refer to $P(x_{1j1}, \ldots, x_{njn})$ as a *joint* parameter and $P(x_{ij})$ as a *marginal* parameter of the corresponding PDM over V.

For any three disjoint subsets of variables W, U and Z in V, subsets W and U are called *conditionally independent* given Z, if

$$P(W \mid U, Z) = P(W \mid Z)$$

for all possible values in W, U and Z such that $P(U, Z) > 0$. Conditional independence signifies the dependence mediated by Z. This allows the dependence among $W \cup U \cup Z$ to be modeled over subsets $W \cup Z$ and $U \cup Z$ separately. Conditional independence is the key property explored through graphical models.

Subsets W and U are said to be *marginally independent* (sometimes referred to as *unconditionally independent*) if

$$P(W \mid U) = P(W)$$

for all possible values W and U such that $P(U) > 0$. When two subsets of variables are marginally independent, there is no dependence between them. Hence, each subset can be modeled independently without losing information.

If each variable x_i in a subset A is marginally independent of $A \setminus \{x_i\}$, the variables in A are said to be *marginally independent*. The following proposition reveals a useful property called *factorization* when this is the case.

- **Proposition 1:** If each variable x_i in a subset A is marginally independent of $A \setminus \{x_i\}$ then

$$P(A) = \prod_{x_i \in A} P(x_i).$$

Variables in a subset A are called *generally dependent*, if $P(B \mid A \setminus B) \neq P(B)$ for every proper subset $B \subset A$. If a subset of variables is generally dependent, its proper subsets cannot be modeled independently without losing information. A generally dependent subset of variables, however, may display conditional independence within the subset. For example, consider $A=\{x_1, x_2, x_3\}$. If $P(x_1, x_2 \mid x_3) = P(x_1, x_2)$, i.e., $\{x_1, x_2\}$ and x_3 are marginally independent, then A is *not* generally dependent. On the other hand, if

$$P(x_1, x_2 \mid x_3) \neq P(x_1, x_2),\ P(x_2, x_3 \mid x_1) \neq P(x_2, x_3),\ P(x_3, x_1 \mid x_2) \neq P(x_3, x_1),$$

then A *is* generally dependent.

Variables in A are *collectively dependent* if, for each proper subset $B \subset A$, there exists no proper subset $C \subset A \setminus B$ that satisfies $P(B \mid A \setminus B) = P(B \mid C)$. Collective dependence prevents conditional independence and modeling through proper subsets of variables. Table 1 shows the JPD over a set of variables $V=\{x_1, x_2, x, x_4\}$. The four variables are collectively dependent; for example,

$$P(x_{1,1'} \mid x_{2.0,} x_{3',1,} x_{4',0}) = 0.257$$

and

$$P(x_{1,1'} \mid x_{2.0,} x_{3',1}) = P(x_{1,1'} \mid x_{2.0,} x_{4',0}) = P(x_{1,1'} \mid x_{3.0,} x_{4',0}) = 0.3.$$

MAIN THRUST

Pseudo-Independent (PI) Models

A *pseudo-independent* (PI) model is a PDM where proper subsets of a set of collectively dependent variables display marginal independence (Xiang, Wong & Cercone, 1997). The basic PI model is a full PI model:

- **Definition 2 (Full PI model):** A PDM over a set V ($|V| \geq 3$) of variables is a *full PI model*, if the following properties (called axioms of full PI models) hold:

  (S_I) Variables in each proper subset of V are marginally independent.

  (S_{II}) Variables in V are collectively dependent.

Table 1 shows the JPD of a binary full PI model, where $V = \{x_1, x_2, x_3, x_4\}$. Its marginal parameters are

$$P(x_{1,0}) = 0.7,\ P(x_{2,0}) = 0.6,\ P(x_{3,0}) = 0.35,\ P(x_{4,0}) = 0.45.$$

Any subset of three variables are marginally independent; for example,

$$P(x_{1,1'} x_{2.0,} x_{3',1}) = P(x_{1,1'}) P(x_{2,0}) P(x_{3',1}) = 0.117.$$

The four variables are collectively dependent as explained previously.

Table 2 is the JPD of the color model given earlier, where $V = \{x_1, x_2, x_3\}$. The marginal independence can be verified by

$$P(x_1{=}red) = P(x_2{=}red) = P(x_3{=}red) = 0.5,$$
$$P(x_1{=}red \setminus x_2) = P(x_1{=}red \setminus x_3) = P(x_2{=}red \setminus x_3) = 0.5,$$

and the collective dependence can be seen from $P(x_1{=}red \setminus x_2{=}red, x_3{=}red) = 1$.

By relaxing condition (S_I) on marginal independence, full PI models are generalized into partial PI models, which are defined through marginally independent partition (Xiang, Hu, Cercone & Hamilton, 2000) introduced in the following:

- **Definition 3** (**Marginally Independent Partition**): Let V ($|V| \geq 3$) be a set of variables, and $B = \{B^1, \ldots, B^m\}$ ($m \geq 2$) be a partition of V. B is a marginally independent partition if, for every subset $A = \{x_i^k \mid x_i^k \in B^k, k = 1,\ldots,m\}$, variables in A are marginally independent. Each block B^i is called a *marginally independent block*.

Intuitively, a marginally independent partition groups variables in V into m blocks. If one forms a subset A by taking one element from each block, then variables in A are marginally independent. Unlike full PI models, in a partial PI model, it is not necessary that every proper subset is marginally independent. Instead, that requirement is replaced with the marginally independent partition.

- **Definition 4 (Partial PI Model):** A PDM over a set V ($|V| \geq 3$) of variables is a partial PI model, if the following properties (called axioms of partial PI models) hold:

  (S_I') V can be partitioned into two or more marginally independent blocks.