Pseudo–Independent Models and Decision Theoretic Knowledge Discovery

Yang Xiang
University of Guelph, Canada

INTRODUCTION

Graphical models such as Bayesian networks (BNs) (Pearl, 1988; Jensen & Nielsen, 2007) and decomposable Markov networks (DMNs) (Xiang, Wong., & Cercone, 1997) have been widely applied to probabilistic reasoning in intelligent systems. Knowledge representation using such models for a simple problem domain is illustrated in Figure 1: Virus can damage computer files and so can a power glitch. Power glitch also causes a VCR to reset. Links and lack of them convey dependency and independency relations among these variables and the strength of each link is quantified by a probability distribution. The networks are useful for inferring whether the computer has virus after checking files and VCR. This chapter considers how to discover them from data.

Discovery of graphical models (Neapolitan, 2004) by testing all alternatives is intractable. Hence, heuristic search are commonly applied (Cooper & Herskovits, 1992; Spirtes, Glymour, & Scheines, 1993; Lam & Bacchus, 1994; Heckerman, Geiger, & Chickering, 1995; Friedman, Geiger, & Goldszmidt, 1997; Xiang, Wong, & Cercone, 1997). All heuristics make simplifying assumptions about the unknown data-generating models. These assumptions preclude certain models to gain efficiency. Often assumptions and models they exclude are not explicitly stated. Users of such heuristics may suffer from such exclusion without even knowing. This chapter e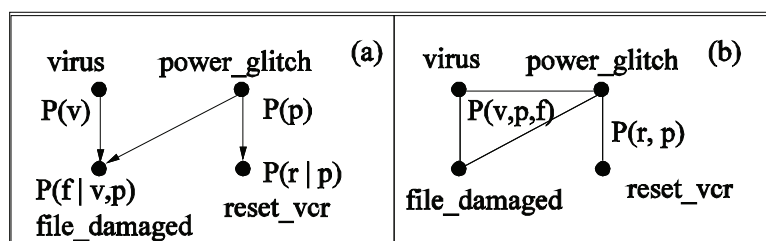xamines assumptions underlying common heuristics and their consequences to graphical model discovery. A decision theoretic strategy for choosing heuristics is introduced that can take into account a full range of consequences (including efficiency in discovery, efficiency in inference using the discovered model, and cost of inference with an incorrectly discovered model) and resolve the above issue.

BACKGROUND

A graphical model encodes probabilistic knowledge about a problem domain concisely (Pearl, 1988; Jensen & Nielsen, 2007). Figure 1 illustrates a BN in (a) and a DMN in (b). Each node corresponds to a binary variable. The graph encodes dependence assumptions among these variables, e.g., that f is directly dependent on v and p, but is independent of r once the value of p is observed. Each node in the BN is assigned a conditional probability distribution (CPD) conditioned on its parent nodes, e.g., $P(f \mid v, p)$ to quantify the uncertain dependency. The joint probability distribution (JPD) for the BN is uniquely defined by the product $P(v, p, f, r) = P(f \mid v, p) \, P(r \mid p) \, P(v) \, P(p)$. The DMN has two groups of nodes that are maximally pairwise connected, called *cliques*. Each is assigned a probability distribution, e.g., $\{v, p, f\}$ is assigned $P(v, p, f)$. The JPD for the DMN is $P(v, p, f) \, P(r, p) / P(p)$.

When discovering such models from data, it is important that the dependence and independence relations

Figure 1. (a) An example BN (b) A corresponding DMN

expressed by the graph approximate true relations of the unknown data-generating model. How accurately can a heuristics do so depends on its underlying assumptions.

To analyze assumptions underlying common heuristics, we introduce key concepts for describing dependence relations among domain variables in this section. Let V be a set of discrete variables $\{x_1, \ldots, x_n\}$. Each x_i has a finite space $S_{xi} = \{x_{i,j} \mid 1 \leq j \leq D_i\}$. When there is no confusion, we write $x_{i,j}$ as x_{ij}. The space of a set $X \subseteq V$ of variables is the Cartesian product $S_X = \prod_{x_i \in X} S_i$. Each element in S_X is a configuration of X, denoted by $x = (x_1, \ldots, x_n)$. A probability distribution $P(X)$ specifies the probability $P(x) = P(x_1, \ldots, x_n)$ for each x. $P(V)$ is the *JPD* and $P(X)$ ($X \subset V$) is a *marginal distribution*. A probabilistic domain model (PDM) over V defines $P(X)$ for every $X \subseteq V$.

For disjoint subsets W, U and Z of V, W and U are *conditionally independent* given Z, if $P(w \mid u, z) = P(w \mid z)$ for all configurations such that $P(u, z) > 0$. The condition is also denoted $P(W \mid U, Z) = P(W \mid Z)$. It allows modeling of dependency within $W \cup U \cup Z$ through *overlapping* subsets $W \cup Z$ and $U \cup Z$.

W and U are *marginally independent* if $P(W \mid U) = P(W)$ holds whenever $P(U) > 0$. The condition allows dependency within $W \cup U$ to be modeled over *disjoint* subsets. If each variable x_i in a subset X is marginally independent of $X \setminus \{x_i\}$, then variables in X are *marginally independent*.

Variables in a subset X are *generally dependent* if $P(Y \mid X \setminus Y) \neq P(Y)$ for every $Y \subset X$. For instance, $X = \{x_1, x_2, x_3\}$ is *not* generally dependent if $P(x_1, x_2 \mid x_3) = P(x_1, x_2)$. It is generally dependent if $P(x_1, x_2 \mid x_3) \neq P(x_1, x_2)$, $P(x_2, x_3 \mid x_1) \neq P(x_2, x_3)$ and $P(x_3, x_1 \mid x_2) \neq P(x_3, x_1)$. Dependency within X cannot be modeled over disjoint subsets but may through overlapping subsets, due to conditional independence in X.

Variables in X are *collectively dependent* if, for each proper subset $Y \subset X$, there exists no proper subset $Z \subset X \setminus Y$ that satisfies $P(Y \mid X \setminus Y) = P(Y \mid Z)$. Collective dependence prevents modeling through overlapping subsets and is illustrated in the next section.

MAIN THRUST OF THE CHAPTER

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). Common heuristics often fail in learning a PI model (Xiang, Wong., & Cercone, 1996). Before analyzing how assumptions underlying common heuristics cause such failure, we introduce PI models below. PI models can be classified into three types: *full, partial, and embedded*. The basic PI model is a full PI model.

Definition 1. A PDM over a set V ($|V| \geq 3$) of variables is a *full PI model* if the following hold:

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

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

Example 1 Patient of a chronicle disease changes the health state (denoted by variable s) daily between stable ($s = t$) and unstable ($s = u$). Patient suffers badly in an unstable day unless treated in the morning, at which time no indicator of the state is detectable. However, if treated at the onset of a stable day, the day is spoiled due to side effect. From historical data, patient's states in four consecutive days observe the estimated distribution in Table 1.

The state in each day is uniformly distributed, i.e., $P(s_i = t) = 0.5$ where $1 \leq i \leq 4$. The state of each day is marginally independent of that of the previous day, i.e., $P(s_i = t \mid s_{i-1}) = 0.5$ where $2 \leq i \leq 4$. It is marginally independent of that of the previous two days, i.e., $P(s_i = t \mid s_{i-1}, s_{i-2}) = 0.5$ where $3 \leq i \leq 4$. However, states of four days are collectively dependent, e.g., $P(s_4 = u \mid s_3 = u, s_2 = t, s_1 = t) = 1$, which allows the state of the last day to be predicted from states of previous three days. Hence, the patient's states form a full PI model.

By relaxing condition *(S_I)*, full PI models are generalized into partial PI models defined through *marginally independent partition* (Xiang, Hu, Cercone, & Hamilton, 2000):

P

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/pseudo-independent-models-decision-theoretic/11037

Related Content

Knowledge Discovery in Databases with Diversity of Data Types

QingXiang Wu, Martin McGinnity, Girijesh Prasadand David Bell (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1117-1123).*
www.irma-international.org/chapter/knowledge-discovery-databases-diversity-data/10961

Evolutionary Computation and Genetic Algorithms

William H. Hsu (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 817-822).*
www.irma-international.org/chapter/evolutionary-computation-genetic-algorithms/10914

Secure Computation for Privacy Preserving Data Mining

Yehuda Lindell (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1747-1752).*
www.irma-international.org/chapter/secure-computation-privacy-preserving-data/11054

Database Security and Statistical Database Security

Edgar R. Weippl (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 610-616).*
www.irma-international.org/chapter/database-security-statistical-database-security/10884

Data Mining Applications in the Hospitality Industry

Soo Kim (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 406-410).*
www.irma-international.org/chapter/data-mining-applications-hospitality-industry/10852