Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence, Cambridge, MA, July 27-29, 1990.

Minimum Error Tree Decomposition

Lifu Liu and David C. Wilkins
Department of Computer Science
University of Illinois
405 North Mathews Ave
Urbana, IL 61801

Xiaoxin Ying and Zhaoqi Bian
Department of Automation
Tsinghua University
Beijing,100084,P.R.China

Abstract

This paper describes a generalization of previous methods for constructing tree-structured belief network with hidden variables. The major new feature of the described method is the ability to produce a tree decomposition even when there are errors in the correlation data among the input variables. This is an important extension of existing methods since the correlational coefficients usually cannot be measured with precision. The technique involves using a greedy search algorithm that locally minimizes an error function.

I. Introduction

In belief network reasoning and learning causal structure, it is desired to decompose distribution $P(x_1, \cdots, x_n)$ of n binary stochastic variables into marginals of $n+1$ variables $P(x_1, \cdots, x_n, w)$, in order to render $x_1, \cdots, x_n$ conditionally independent given w (J. Pearl 1986).

This decomposition problem is very difficult to solve directly (P.Lazarfeld 1966). Pearl's solution approach to the problem was to replace the hidden variable w with a group of hidden variables $w_1, \cdots, w_m, 1 \leq m \leq n-2$, and then construct a tree that has n leaves corresponding to the stochastic variables $x_1, \cdots, x_n$ and has m internal hidden nodes $w_1, \cdots, w_m$ (Pearl, 1986). Thus the $P(x_1, \cdots, x_n)$ is represented as the margin of the correlation distribution of the nodes in the tree,

$$P(x_1, \cdots, x_n, w_1, \cdots, w_m), 1 \leq m \leq n-2.$$

The first step of Pearl's method is to define the four allowable topologies of quadruplets, where each quadruplet has four leaves i, j, k, l, and two internal nodes w_1 and w_2, and each of the internal nodes is connected to a pair of the leaves. Given this fundamental tree substructure, Pearl's method defines a sufficient condition for connecting four leaves into a quadruplet. This sufficient condition is then applied repeatedly to construct a tree.

In this paper, we extend the excellent approach to Pearl by defining a weaker sufficient condition for connecting four leaves into a quadruplet. This weaker sufficient condition is satisfied in situations where the input information is noisy or incomplete, which is typically the case in actual applications.

The sufficient condition for the topology of quadruplet used by Pearl is

$$\rho_{ik}\rho_{jl} = \rho_{il}\rho_{jk}, \tag{1}$$

given that it is known that $P(x_1, \cdots, x_n)$ can be written as the marginal probability of an $(n+m)$ variable distribution $P_s(x_1, \cdots, x_n, w_1, \cdots, w_m)$ in which $x_1, \cdots, x_n$ are conditionally independent given $w_1, \cdots, w_n$,i.e.,

$$P_s(x_1, \cdots, x_n, w_1, \cdots, w_m) = \tag{2}$$

$$\prod_{i=1}^{n} P_s(x_i|w_j)P_s(w_j|w_k) \cdots P(w_m) \tag{3}$$

where

$$\rho_{ij} = \frac{(p_{ij} - p_i p_j)}{(p_i(1-p_i))^{1/2}(p_j(1-p_j))^{1/2}},$$

$$p_i = P(x_i = 1),$$

$$p_{ij} = P(x_i = 1, x_j = 1).$$

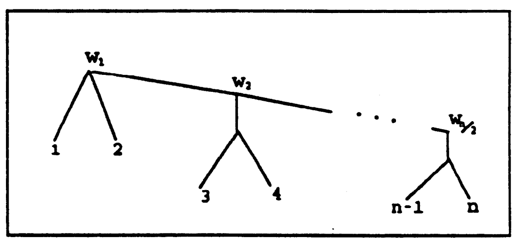

Figure 1. The type of tree structure that is produced by repeated application of equation (1).

An ideal decomposition will satisfy the conditions (1) for all nodes connected to internal nodes. Unfortunately, this is often not the case, and so the decomposition algorithm will terminate without producing a solution. An approach that circumvents this problem is to minimize the error $|\rho_{ik}\rho_{jl} - \rho_{il}\rho_{jk}|$.

tree branches below a node if the error at the node is greater than the current minimum error.

It can be easily deduced that this procedure requires exponential time, and needs an additional step to determine the connecting topology once the node quadruplets are determined. This ends our presentation of our earlier approach, which was presented to illustrate some tions are more likely in a decomposed tree than the others. Examine two trees in Figure 4; correlations between pairs (1,2) and (3,4) in (a) are thought to be more "reliable" than that in (b). From a mathematical point of view, correlations in (a)

$$\rho_{14} = \rho_{1w_1}\rho_{w_1w_2}\rho_{w_24}$$

$$\rho_{24} = \rho_{2w_1}\rho_{w_1w_2}\rho_{w_24}$$