International Journal of Approximate Reasoning 52 (2011) 641–657

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Inference in hybrid Bayesian networks using mixtures of polynomials

Prakash P. Shenoy *, James C. West

University of Kansas School of Business, 1300 Sunnyside Ave., Summerfield Hall, Lawrence, KS 66045-7601, USA

ARTICLE INFO

Article history:
Available online 26 September 2010

Keywords:
Hybrid Bayesian networks
Inference in hybrid Bayesian networks
Shenoy–Shafer architecture
Extended Shenoy–Shafer architecture
Mixtures of polynomials
Mixtures of truncated exponentials

ABSTRACT

The main goal of this paper is to describe inference in hybrid Bayesian networks (BNs) using mixture of polynomials (MOP) approximations of probability density functions (PDFs). Hybrid BNs contain a mix of discrete, continuous, and conditionally deterministic random variables. The conditionals for continuous variables are typically described by conditional PDFs. A major hurdle in making inference in hybrid BNs is marginalization of continuous variables, which involves integrating combinations of conditional PDFs. In this paper, we suggest the use of MOP approximations of PDFs, which are similar in spirit to using mixtures of truncated exponentials (MTEs) approximations. MOP functions can be easily integrated, and are closed under combination and marginalization. This enables us to propagate MOP potentials in the extended Shenoy–Shafer architecture for inference in hybrid BNs that can include deterministic variables. MOP approximations have several advantages over MTE approximations of PDFs. They are easier to find, even for multi-dimensional conditional PDFs, and are applicable for a larger class of deterministic functions in hybrid BNs.

1. Introduction

Bayesian networks (BNs) and influence diagrams (IDs) were invented in the mid 80s (see e.g., [26,12]) to represent and reason with large multivariate discrete probability models and decision problems, respectively. Several efficient algorithms exist to compute exact marginals of posterior distributions for discrete BNs (see e.g., [18,35,13]) and to solve discrete IDs exactly (see e.g., [25,32,33,14]).

Hybrid Bayesian networks contain a mix of discrete and continuous variables. A continuous variable is said to be *deterministic* if its conditional distributions have zero variances. The conditional distributions of deterministic variables are typically described by equations that describe the deterministic variable as a function of its continuous parents. Deterministic variables pose a problem in inference since the joint density of all continuous variables does not exist. Shenoy and West [36] describe an extension of the Shenoy–Shafer architecture [35] to enable inference in hybrid BNs with deterministic variables.

The state of the art exact algorithm for mixtures of Gaussians hybrid BNs is the Lauritzen and Jensen [19] algorithm implemented with Madsen's [22] lazy propagation technique. This requires the conditional PDFs of continuous variables to be conditional linear Gaussians (CLGs), and that discrete variables do not have continuous parents. Marginals of multivariate normal distributions can be found easily without the need for integration. The disadvantages are that in the inference process, continuous variables have to be marginalized before discrete ones. In some problems, this restriction can lead to large cliques [21].

If a BN has discrete variables with continuous parents, Murphy [24] uses a variational approach to approximate the product of the potentials associated with a discrete variable and its parents with a CLG distribution. Lerner [20] uses a numerical

* Corresponding author. Tel.: +1 785 864 7551; fax: +1 785 864 5328.
E-mail addresses: pshenoy@ku.edu (P.P. Shenoy), cully@ku.edu (J.C. West).

0888-613X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2010.09.003

integration technique called Gaussian quadrature to approximate non-CLG distributions with CLG distributions, and this same technique can be used to approximate the product of potentials associated with a discrete variable and its continuous parents. Murphy's and Lerner's approach is then embedded in the Lauritzen and Jensen [19] algorithm to solve the resulting mixtures of Gaussians BN.

Shenoy [34] proposes approximating non-CLG distributions by mixtures of Gaussians using a non-linear optimization technique, and using arc reversals to ensure discrete variables do not have continuous parents. The resulting mixture of Gaussians BN is then solved using the Lauritzen and Jensen [19] algorithm.

Moral et al. [23] propose approximating PDFs by mixtures of truncated exponentials (MTEs), which are easy to integrate in closed form. Since the family of mixtures of truncated exponentials is closed under combination and marginalization, the Shenoy and Shafer [35] architecture can be used to solve an MTE BN. Cobb and Shenoy [5] and Cobb et al. [6] propose using a non-linear optimization technique for finding MTE approximations for several commonly used one-dimensional distributions. Cobb and Shenoy [3,4] extend this approach to BNs with linear and non-linear deterministic variables. In the latter case, they approximate non-linear deterministic functions by piecewise linear ones. Rumi and Salmeron [28] describe approximate probability propagation with MTE approximations that have only two exponential terms in each piece. Romero et al. [27] describe learning MTE potentials from data, and Langseth et al. [17] investigate the use of MTE approximations where the coefficients are restricted to integers.

In this paper, we propose using mixture of polynomials (MOP) approximations of PDFs. Mixtures of polynomials are widely used in many domains including computer graphics, font design, approximation theory, and numerical analysis. They were first studied by Schoenberg [30]. When the MOP functions are continuous, they are referred to as *polynomial splines* [8,31]. The use of splines to approximate PDFs was initially suggested by Curds [7]. For our purposes, continuity is not an essential requirement, and we will restrict our analysis to piecewise polynomial approximations of PDFs.

Using MOP is similar in spirit to using MTEs. MOP functions can be easily integrated, and they are closed under combination and marginalization. Thus, the extended Shenoy–Shafer architecture [36] can be used to make inferences in BN with deterministic variables. However, there are several advantages of MOP functions over MTEs.

First, we can find MOP approximations of differentiable PDFs easily by using the Taylor series approximations. Finding MTE approximations as suggested by Cobb et al. [6] necessitates solving non-linear optimization problems, which is not as easy a task as it involves navigating among local optimal solutions.

Second, for the case of conditional PDFs with several parents, finding a good MTE approximation can be extremely difficult as it involves solving a non-linear optimization problem in a high-dimensional space for each piece. The Taylor series expansion can also be used for finding MOP approximations of conditional PDFs. In Section 2, we describe a MOP approximation for a 2-dimensional CLG distribution.

Third, if a hybrid BN contains deterministic functions, then the MTE approach can be used directly only for linear deterministic functions. By directly, we mean without approximating a non-linear deterministic function by a piecewise linear one. This is because the MTE functions are not closed under transformations needed for non-linear deterministic functions. MOP functions are closed under a larger family of deterministic functions including linear functions and quotients. This enables propagation in a bigger family of hybrid BNs than is possible using MTEs.

An outline of the remainder of the paper is as follows. In Section 2, we define MOP functions and describe how one can find MOP approximations with illustration for the univariate normal distribution, chi-square distribution, and for a two-dimensional CLG distribution. In Section 3, we sketch the extended Shenoy–Shafer architecture for inference in hybrid BNs with deterministic variables. In Section 4, we solve three small examples designed to demonstrate the feasibility of using MOP approximations with non-linear deterministic functions. Finally, in Section 5, we end with a summary and discussion of some of the challenges associated with MOP approximations.

2. Mixtures of polynomials approximations

In this section, we describe MOP functions and some methods for finding MOP approximations of PDFs. We illustrate our method for the normal distribution, the chi-square distribution, and the CLG distribution in two dimensions.

A one-dimensional function $f : \mathbb{R} \to \mathbb{R}$ is said to be a *mixture of polynomials* (MOP) function if it is a piecewise function of the form:

$$\begin{aligned} f(x) &= a_{0i} + a_{1i}x + a_{2i}x^2 + \cdots + a_{ni}x^n, \quad \text{for } x \in A_i, \quad i = 1, \ldots, k, \text{ and} \\ &= 0, \quad \text{otherwise,} \end{aligned} \tag{2.1}$$

where $A_1, \ldots, A_k$ are disjoint intervals in $\mathbb{R}$ that do not depend on x, and $a_{0i}, \ldots, a_{ni}$ are constants for all i. We will say that f is a *k-piece* (ignoring the 0 piece), and *n-degree* (assuming $a_{ni} \neq 0$ for some i) MOP function.

The main motivation for defining MOP functions is that such functions are easy to integrate in closed form, and that they are closed under multiplication and integration. They are also closed under differentiation and addition.

An m-dimensional function $f : \mathbb{R}^m \to \mathbb{R}$ is said to be a MOP function if

$$f(x_1, \ldots, x_m) = f_1(x_1) \cdot f_2(x_2) \cdots f_m(x_m), \tag{2.2}$$