Decision Support Systems 53 (2012) 580–590

Contents lists available at SciVerse ScienceDirect

Decision Support Systems

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

Answering queries in hybrid Bayesian networks using importance sampling

Antonio Fernández *, Rafael Rumí, Antonio Salmerón

Dept. Statistics and Applied Mathematics, University of Almería, La Cañada de San Urbano s/n, 04120 Almería, Spain

ARTICLE INFO

Article history:
Received 25 February 2011
Received in revised form 1 February 2012
Accepted 24 March 2012
Available online 3 April 2012

2000 MSC:
6505
68T37

Keywords:
Bayesian networks
Probabilistic reasoning
Importance sampling
Mixtures of truncated exponentials

ABSTRACT

In this paper we propose an algorithm for answering queries in hybrid Bayesian networks where the underlying probability distribution is of class MTE (mixture of truncated exponentials). The algorithm is based on importance sampling simulation. We show how, like existing importance sampling algorithms for discrete networks, it is able to provide answers to multiple queries simultaneously using a single sample. The behaviour of the new algorithm is experimentally tested and compared with previous methods existing in the literature.

1. Introduction

Bayesian networks [17,36] have become a popular tool for representing uncertainty in decision support systems. A review of recent literature shows the variety of applications in which they have been successfully used [1,10,24,35,45]. One of the main reasons for using them as the inference engine in a decision support system is that efficient reasoning algorithms can be designed, taking advantage of their structure [2,3,16,44,43,30,29].

Most of the methodological development around Bayesian networks has concentrated on the case in which all the variables involved are qualitative or discrete. However, decision support systems usually have to operate in domains described in terms of both discrete and continuous variables simultaneously. In such scenarios, there is always the possibility of discretising the continuous variables [20,34], in order to be able to use methods designed for discrete variables. But such a solution in general conveys a loss of information.

Continuous and discrete variables can be handled simultaneously, with no need to discretise, in the so-called *hybrid Bayesian networks*. The first advances in this field came along with the definition of the Conditional Gaussian (CG) model [25,26,28]. The limitations of this approach are the assumption of normality over the continuous variables, and also the fact that dependencies of discrete variables conditional on continuous ones, are not allowed. This structural restriction is overcome in the *augmented Conditional Linear Gaussian (CLG) networks*, where discrete nodes are allowed to have continuous parents, by representing their conditional distributions as *softmax* functions [27]. However this model also relies on the normality assumption. Furthermore, exact inference is not possible in augmented CLG networks, and the solution proposed in [27] is based on a Gaussian approximation of the product of the Gaussian and softmax functions, which provides exact marginals for the discrete variables and also is able to obtain exact values only for the first and second order moments of the distribution of the continuous variables.

A more general proposal is based on the use of mixtures of truncated exponentials (MTEs), which do not impose any restriction and also do not rely on the normality assumption [31]. This model has been successfully applied to decision problems [6]. An important feature of MTEs is that they are compatible with efficient exact inference algorithms like, for instance, the Shenoy–Shafer architecture [44] and the variable elimination scheme [49]. As MTEs are able to approximate a wide variety of probability distributions [7], they can be used as a general framework for carrying out inference in hybrid Bayesian networks, just by approximating each conditional distribution in the network by an MTE and then using an exact inference algorithm. This approach has been analysed in [22], by solving a network involving Logistic and Gaussian distributions using MTEs, variational approximations [18], discretisation [33] and Markov Chain Monte Carlo [12].

A recent approach, similar in essence to MTEs, is based on representing the distribution in a hybrid Bayesian network as a *Mixture of Polynomials* (*MOPs*) [42]. Both MTEs and MOPs have been generalised in a global framework for representing hybrid Bayesian networks, called *Mixtures of Truncated Basis Functions* (*MoTBFs*) [23]. However, even though MOPs have some advantages over MTEs, specially the

* Corresponding author. Tel.: +34 950214650; fax: +34 950015167.
E-mail addresses: afalvarez@ual.es (A. Fernández), rrumi@ual.es (R. Rumí), antonio.salmeron@ual.es (A. Salmerón).

0167-9236/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2012.03.007

ability of dealing with a wider class of deterministic relationships, so far they lack of an algorithm for learning the models from data, while this issue has been solved for MTEs [38]. Hence, MTEs can be used as an exact model and not only as an approximation of other distributions. In that sense, MTEs behave as a nonparametric model, where no assumption is made about the underlying distribution.

Even though Bayesian networks allow efficient inference algorithms to operate over them, it is known that exact probabilistic inference is an NP-hard problem [8]. Furthermore, approximate probabilistic inference is also an NP-hard problem if a given precision is required [9]. For that reason, approximate algorithms that tradeoff complexity for accuracy have been developed for discrete Bayesian networks. An important class of such approximate algorithms is based on the importance sampling technique, that provides a flexible approach to construct anytime reasoning algorithms [4,13,32,46–48].

Inference in hybrid Bayesian networks with MTEs does not escape from the above mentioned complexity. If the model is learnt from a database using the algorithm in [38], it can be too complex if the number of variables is high. But even using the approximations in [7], inference may become unfeasible if the model is complex enough.

With this motivation, in this paper we propose an approximate algorithm for computing fast and accurate answers to precise queries in hybrid Bayesian networks with MTEs. The algorithm is based on importance sampling, and therefore it is an *anytime* algorithm [37] in the sense that the accuracy of its results is proportional to the time it is allowed to use for computing the answer. We show how our proposal outperforms the previous state-of-the-art method for approximate inference with MTEs, introduced in [40].

The rest of the paper is organised as follows. We establish the notation and define some preliminary concepts in Section 2. The problem addressed here is formally posted in Section 3. The core of the methodological contributions is in Section 4, and the details of the algorithm can be found in Section 5. The experimental analysis carried out to test the performance of the algorithm is reported in Section 6. The concluding remarks are given in Section 7.

2. Notation and preliminaries

Formally, a *Bayesian network* is a directed acyclic graph where each node represents a random variable, and the topology of the graph encodes the independence relations among the variables, according to the *d*-separation criterion [36]. Given the independences attached to the graph, the joint distribution is determined giving a probability distribution for each node conditioned on its parents, so that for a Bayesian network with variables $X_1,\ldots,X_n$, the joint distribution factorises as

$$p(x_1,\ldots,x_n) = \prod_{i=1}^{n} p(x_i|pa(x_i)), \tag{1}$$

where $pa(x_i)$ denotes the parents of variable X_i in the network.

We will use uppercase letters to denote random variables, and boldfaced uppercase letters to denote random vectors, e.g. $\mathbf{X}=\{X_1,\ldots,X_n\}$, and its domain will be written as $\Omega_{\mathbf{X}}$. By lowercase letters x (or $\mathbf{x}$) we denote some element of Ω_X (or $\Omega_{\mathbf{X}}$).

We are interested in *hybrid* Bayesian networks, which are defined for a set of variables $\mathbf{X}$ that contains discrete and continuous variables. Throughout this paper we will assume that $\mathbf{X}=\mathbf{Y}\cup\mathbf{Z}$, being $\mathbf{Y}$ and $\mathbf{Z}$ sets containing only discrete and only continuous variables respectively. We will follow the approach based on mixtures of truncated exponentials [31], in which all the conditional distributions in Eq. (1) are represented as MTE potentials, which are formally defined as follows.

Definition 1. MTE potential

Let $\mathbf{X}$ be a mixed n-dimensional random vector. Let $\mathbf{Y}=(Y_1,\ldots,Y_d)^T$ and $\mathbf{Z}=(Z_1,\ldots,Z_c)^T$ be the discrete and continuous parts of $\mathbf{X}$, respectively, with $c+d=n$. We say that a function $f:\Omega_{\mathbf{X}}\mapsto\mathbb{R}_0^+$ is a mixture of truncated exponentials (MTE) potential if for each fixed value $\mathbf{y}\in\Omega_{\mathbf{Y}}$ of the discrete variables $\mathbf{Y}$, the potential over the continuous variables $\mathbf{Z}$ is defined as:

$$f(\mathbf{z}) = a_0 + \sum_{i=1}^{m} a_i exp\left\{\mathbf{b}_i^T\mathbf{z}\right\}, \tag{2}$$

for all $\mathbf{z}\in\Omega_{\mathbf{Z}}$, where $a_i\in\mathbb{R}$ and $\mathbf{b}_i\in\mathbb{R}^c$, $i=1,\ldots,m$. We also say that f is an MTE potential if there is a partition $D_1,\ldots,D_k$ of $\Omega_{\mathbf{Z}}$ into hypercubes and in each one of them, f is defined as in Eq. (2). An MTE potential is an MTE density if it integrates to 1.

A *conditional MTE density* can be specified by dividing the domain of the conditioning variables and specifying an MTE density for the conditioned variable for each configuration of splits of the conditioning variables. The next is an example of a conditional MTE density.

$$f(y|x) = \begin{cases} 1.26-1.15e^{0.006y} & \text{if } 0.4\le x<5,\ 0\le y<13,\\ 1.18-1.16e^{0.0002y} & \text{if } 0.4\le x<5,\ 13\le y<43,\\ 0.07-0.03e^{-0.4y}+0.0001e^{0.0004y} & \text{if } 5\le x<19,\ 0\le y<5,\\ -0.99+1.03e^{0.001y} & \text{if } 5\le x<19,\ 5\le y<43. \end{cases}$$

Since MTEs are defined into hypercubes, they admit a tree-structured representation in a natural way. Each entire branch in the tree determines one hypercube where the potential is defined, and the function stored in the leaf of a branch is the definition of the potential on it. An example of a tree-structured representation of an MTE potential is shown in Fig. 1.

We use the term *mixed tree* [31] to refer to a tree-structure representation of an MTE potential. A tree $\mathcal{T}$ is a *mixed tree* if: (i) every internal node represents a random variable, (ii) every arc outgoing from a continuous variable Z is labelled with an interval of values of Z, so that the domain of Z is the union of the intervals corresponding to the arcs Z-outgoing, (iii) every discrete variable has a number of outgoing arcs equal to its number of states and (iv) each leaf node contains an MTE potential defined on variables in the path from the root to that leaf.

3. Problem formulation

The goal of this paper is to introduce a method for answering queries in hybrid Bayesian networks with MTEs. We consider a hybrid Bayesian network defined for a set of variables $\mathbf{X}$. A *query* is a question about a probability value for a target variable $W\in\mathbf{X}$ given that the values of some variables $\mathbf{E}\subset\mathbf{X}$ are known. Thus, if we write $\mathbf{X}=(W,\mathbf{Y}^T,\mathbf{Z}^T,\mathbf{E}^T)^T$, where $\mathbf{Y}=(Y_1,\ldots,Y_d)^T$ represents the non-observed discrete variables and $\mathbf{Z}=(Z_1,\ldots,Z_c)^T$ represents the non-observed continuous variables and $\mathbf{E}=(E_1,\ldots,E_k)^T$, then a query about W given that $\mathbf{E}=\mathbf{e}$ is

$$P(a<W<b|\mathbf{E}=\mathbf{e}) = \frac{\int_a^b \left(\sum_{\mathbf{y}\in\mathbf{Y}}\int_{\Omega_{\mathbf{Z}}}\phi(w,\mathbf{y},\mathbf{z},\mathbf{e})d\mathbf{z}\right)dw}{\phi_{\mathbf{E}}(\mathbf{e})} \tag{3}$$

if W is a continuous variable. The function ϕ in Eq. (3) is the joint distribution in the network and $\phi_{\mathbf{E}}$ is its marginal over variables $\mathbf{E}$. Let ϕ_X denote the conditional distribution of any variable X in the network. Then, the joint distribution is defined as

$$\phi(w,\mathbf{y},\mathbf{z},\mathbf{e}) = \phi_W(w|pa(w))\prod_{i=1}^{d}\phi_{Y_i}(y_i|pa(y_i))\prod_{j=1}^{c}\phi_{Z_j}\left(z_j|pa\left(z_j\right)\right)\prod_{l=1}^{k}\phi_{E_l}(e_l|pa(e_l)). \tag{4}$$

Since our goal is to answer a query given a fixed value $\mathbf{e}$ of variables $\mathbf{E}$, we will rather be interested in the restriction of the joint distribution to the knowledge that $\mathbf{E}=\mathbf{e}$. We will replace any symbol ϕ in Eq. (4) by