International Journal of Approximate Reasoning 92 (2018) 155–174

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

On conditional truncated densities Bayesian networks ☆

Santiago Cortijo, Christophe Gonzales *

Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6, UMR 7606, Paris, France

ARTICLE INFO

Article history:
Received 16 October 2016
Received in revised form 4 September 2017
Accepted 2 October 2017
Available online 16 October 2017

Keywords:
Bayesian network
Continuous random variable
Mixed probability distribution
Inference

ABSTRACT

The majority of Bayesian networks learning and inference algorithms rely on the assumption that all random variables are discrete, which is not necessarily the case in real-world problems. In situations where some variables are continuous, a trade-off between the expressive power of the model and the computational complexity of inference has to be done: on one hand, conditional Gaussian models are computationally efficient but they lack expressive power; on the other hand, mixtures of exponentials (MTE), basis functions (MTBF) or polynomials (MOP) are expressive but this comes at the expense of tractability. In this paper, we introduce an alternative model called a ctdBN that lies in between. It is composed of a "discrete" Bayesian network (BN) combined with a set of univariate conditional truncated densities modeling the uncertainty over the continuous random variables given their discrete counterpart resulting from a discretization process. We prove that ctdBNs can approximate (arbitrarily well) any Lipschitz mixed probability distribution. They can therefore be exploited in many practical situations. An efficient inference algorithm is also provided and its computational complexity justifies theoretically why inference computation times in ctdBNs are very close to those in discrete BNs. Experiments confirm the tractability of the model and highlight its expressive power, notably by comparing it with BNs on classification problems and with MTEs and MOPs on marginal distributions estimations.

1. Introduction

For several decades, Bayesian networks (BN) [1] have been successfully exploited for dealing with uncertainties. Their popularity has stimulated the development of many efficient learning and inference algorithms [2–6]. Whilst these algorithms are relatively well understood when they involve only discrete variables, their ability to cope with continuous variables is often unsatisfactory. Dealing with continuous random variables is much more complicated than dealing with discrete ones and one actually has to trade-off between the expressive power of the uncertainty model and the computational complexity of its learning and inference mechanisms. Conditional Gaussian models and their mixing with discrete variables [7–9] lie on one side of the spectrum. They compactly represent multivariate Gaussian distributions. Their inference mechanisms are computationally very efficient but their main drawback is that they lack expressive power. Indeed, although Conditional Linear Gaussian (CLG) models can easily encode conditional independences between random variables, the density functions of their continuous random variables are required to be Normal distributions whose parameters depend linearly on the values of their parents. They are therefore unable to represent models where dependences between

☆ This paper is part of the Virtual special issue on Uncertainty Reasoning, edited by Robert E. Mercer and Salem Benferhat.

* Corresponding author.
E-mail addresses: Santiago.Cortijo@lip6.fr (S. Cortijo), Christophe.Gonzales@lip6.fr (C. Gonzales).

https://doi.org/10.1016/j.ijar.2017.10.007
0888-613X/

the continuous random variables are nonlinear. In addition, they are not very well suited to represent models in which random variables are not distributed w.r.t. Normal distributions. On the other side of the spectrum, there are more expressive models like mixtures of exponentials (MTE) [10–12], mixtures of truncated basis functions (MTBF) [13,14] and mixtures of polynomials (MOP) [15–17]. Those can approximate very well density functions but this comes at the expense of tractability: their exact inference computation times tend to grow exponentially with the number of continuous variables, which makes them unusable when they contain hundreds of random variables.

In this paper, we propose an alternative model that lies in between these two extremes. The key idea is to discretize the random variables, thereby mapping each (continuous) value of their domain into an interval within a *finite* set of intervals. Of course, whenever some discretization is performed, some information about the continuous random variables is lost. But this can be significantly alleviated by modeling the distribution of the continuous values within each discretization interval by a density function which may not necessarily be a uniform distribution (which is the implicit assumption when using a classical discretization). The set of density functions over all the intervals of a continuous variable constitutes its "*conditional truncated density*" given its discretized counterpart. Now, our uncertainty model is a (discrete) BN over the set of discrete and discretized random variables combined with the set of conditional truncated densities assigned to the continuous random variables that were discretized. This model is therefore called "*conditional truncated densities Bayesian network*", or ctdBN for short. It represents compactly mixed probability distributions. The model is derived from the result of an algorithm for learning BNs from datasets containing both discrete and continuous random variables [18].

By assigning conditional truncated densities to continuous variables, our model gains expressive power over a BN in which all continuous variables are discretized. As we show in this paper, this assertion is justified theoretically by the fact that any Lipschitz mixed probability distribution can be (arbitrarily well) approximated by a ctdBN. For inference, the density functions need only be included in the BN as discrete evidence (computed by integrations) over the discretized variables and, then, only a classical inference over discrete variables is needed to complete the process. As, in our model, the density functions are univariate, integrations can be performed efficiently. So the inference times are very close to those of inferences in classical BNs, which makes inference tractable in ctdBNs. The theoretical computational complexity of our inference algorithm supports this assertion. In addition, the experiments performed in the paper also highlight this point as well as the expressive power of the model.

The paper is organized as follows. In the next section, we recall some related works on CLGs, MTEs, MTBFs and MOPs. Then, in Section 3, we present our model, we study theoretically its expressive power, i.e., its capacity to approximate mixed probability distributions, and we propose an inference algorithm as well as its computational complexity. Next, the efficiency and effectiveness of ctdBN's inferences are highlighted through a set of experiments. Finally, a conclusion and some perspectives are provided in the last section.

2. Related works

In the rest of the paper, capital letters (possibly subscripted) refer to random variables and boldface capital letters to sets of variables. To distinguish continuous random variables from discrete ones, we denote by $\mathring{X}_i$ a continuous variable and by X_i a discrete one. Without loss of generality, for any $\mathring{X}_i$, variable X_i represents its discretized counterpart. Throughout the paper, let $\mathbf{X_D} = \{X_1, \dots, X_d\}$ and $\mathring{\mathbf{X}}_\mathbf{C} = \{\mathring{X}_{d+1}, \dots, \mathring{X}_n\}$ denote the set of discrete and continuous random variables respectively. We denote by $\mathcal{X} = \mathbf{X_D} \cup \mathring{\mathbf{X}}_\mathbf{C}$ the set of all random variables. In addition, for any set of indices $\mathbf{I} = \{i_1, \dots, i_k\}$, $\mathbf{X_I}$ denotes the set of random variables $\{X_{i_1}, \dots, X_{i_k}\}$. Finally, for any variable X or set of random variables $\mathbf{Y}$ or $\mathring{\mathbf{Y}}$, let Ω_X (resp. $\Omega_\mathbf{Y}$ or $\Omega_{\mathring{\mathbf{Y}}}$) denote the domain of X (resp. $\mathbf{Y}$ or $\mathring{\mathbf{Y}}$).

As mentioned in the introduction, a Conditional Linear Gaussian (CLG) model represents a mixed probability distribution [7]. Like in a BN, in the CLG model, to each discrete variable X_i in $\mathbf{X_D}$ is assigned its conditional probability table (CPT) $P(X_i|\mathbf{Pa}(X_i))$ given its parents (the latter are all discrete). In addition, to each continuous variable $\mathring{X}_i \in \mathring{\mathbf{X}}_\mathbf{C}$ is assigned the conditional distribution:

$$f(\mathring{x}_i|\mathbf{X_{D_i}} = \mathbf{x_{D_i}}, \mathring{\mathbf{X}}_\mathbf{C_i} = \mathring{\mathbf{x}}_\mathbf{C_i}) = \mathcal{N}(\mathring{x}_i|\alpha(\mathbf{x_{D_i}}) + \beta(\mathbf{x_{D_i}})^T \mathring{\mathbf{x}}_\mathbf{C_i}, \sigma(\mathbf{x_{D_i}})),$$

where $\mathbf{X_{D_i}}$ and $\mathring{\mathbf{X}}_\mathbf{C_i}$ are the set of discrete and continuous parents of $\mathring{X}_i$ respectively. $\alpha(\mathbf{x_{D_i}})$ and $\beta(\mathbf{x_{D_i}})$ are the coefficients of a linear regression model of $\mathring{X}_i$ given its continuous parents. These coefficients depend on the values $\mathbf{x_{D_i}}$ of the discrete parents. The product of all the CPTs and the conditional distributions represent the joint mixed distribution over $\mathcal{X}$. Our ctdBN model shares some similarities with CLGs: it represents mixed probability distributions using a DAG whose nodes represent random variables, the parents of the discrete ones being also necessarily discrete. The main difference between CLGs and ctdBNs lies in the conditional density functions assigned to the continuous random variables. Unlike CLGs, in ctdBNs, they are not limited to normal distributions and any conditional truncated density function can be used. In this sense, ctdBNs are more general than CLGs. The dependences between the continuous variables are also different: in CLGs, those are necessarily linear (the coefficients of the mean vector result from a linear regression) whereas, in ctdBNs, linearity is not required. Of course, nonlinearity can be approximated by piecewise linear functions and can therefore be taken into account in CLGs introducing latent variables and using deterministic relationships, like in [19]. In ctdBNs, this is taken into account directly through the relationships between the discretized random variables. There is no need to introduce latent variables, which complexifies the learning of the model. Finally, in CLGs, the mean vector of the normal distribution assigned