Expert Systems with Applications 38 (2011) 12409–12414

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Evaluating the difference between graph structures in Gaussian Bayesian networks

Miguel Ángel Gómez-Villegas [a], Paloma Main [a,*], Hilario Navarro [b], Rosario Susi [c]

[a] Dpto. Estadística e I.O., Fac. Ciencias Matemáticas, Univ. Complutense de Madrid, 28040 Madrid, Spain
[b] Dpto. de Estadística, I.O. y Cálc. Numérico, Fac. Ciencias, UNED, 28040 Madrid, Spain
[c] Dpto. Estadística e I.O. III, E.U. Estadística, Univ. Complutense de Madrid, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Gaussian Bayesian networks
Conditional specification
Sensitivity analysis
Kullback–Leibler divergence measure

ABSTRACT

In this work, we evaluate the sensitivity of Gaussian Bayesian networks to perturbations or uncertainties in the regression coefficients of the network arcs and the conditional distributions of the variables. The Kullback–Leibler divergence measure is used to compare the original network to its perturbation. By setting the regression coefficients to zero or non-zero values, the proposed method can remove or add arcs, making it possible to compare different network structures. The methodology is implemented with some case studies.

1. Introduction

A Bayesian network (BN) is a probabilistic model of causal interactions between a set of variables, where the joint probability distribution is described in graphical terms. Probabilistic networks have become an increasingly popular paradigm for reasoning through uncertain, complex models in a variety of situations, including AI, medical diagnosis and data mining (Kjærulff & Madsen, 2008).

This model consists of two parts: one qualitative and the other quantitative. Its qualitative aspect is a directed, acyclic graph (DAG), with nodes and arcs that represent a set of variables and their relationships respectively. Based on the dependence structure depicted in the graph, the joint distribution of the variables can be factorized in terms of the univariate conditional distributions of each variable given its parents in the DAG. These distributions constitute the quantitative portion of the model.

Building a BN is a difficult task, because all of the individual distributions and relationships between variables need to be correctly specified. Expert knowledge is essential to fix the dependence structure among variables of the network and to determine a large set of parameters. Databases can aid the process, but provide incomplete data and only partial knowledge of the domain. Thus, any assessments obtained using only databases are inevitably inaccurate (van der Gaag, Renooij, & Coupe, 2007).

The present research is restricted to a subclass of BNs known as Gaussian Bayesian networks (GBNs). The quantitative portion of a GBN consists of a univariate normal distribution for each variable given its parents in the DAG. Also, the joint probability distribution of the model is constrained to be a multivariate normal distribution.

For each variable X_i, the experts have to provide its mean, the regression coefficients between X_i and each parent $X_j \in pa(X_i) \subset \{X_1, \ldots, X_{i-1}\}$, and the conditional variance of X_i given its parents. This specification is easy for experts, because they only have to describe univariate distributions. Moreover, the arcs in the DAG can be expressed in terms of the regression coefficients.

Our interest in this paper is the sensitivity of GBNs defined by these parameters. This subject has not been frequently treated in the literature, because sensitivity analyses usually perturb the joint parameters instead of the conditional parameters. However, it is easy to model the presence or absence of arcs by adopting regression coefficients different from or equal to zero. Thus, it is also possible to study the effect of changes in the qualitative part of the network. An objective evaluation of this effect may also reveal that a simpler DAG structure yields equivalent results.

In Section 1 we define GBNs, present some general concepts, and introduce a working example. In Section 2 we describe the methodology used to study the sensitivity of the GBN and calculate the sensitivity of our example. In Section 3 we vary the network structure and present a metrology example: the calibration of an electronic level using a sine table. The paper ends with some conclusions and suggestions for further research.

2. General concepts

Throughout this work, random variables will be denoted by capital letters. Moreover, in the multidimensional case, boldface characters will be used.

A BN is defined by a pair $(\mathcal{G}, \mathcal{P})$, where $\mathcal{G}$ is the DAG with nodes representing variables and arcs showing the dependence structure. $\mathcal{P}$ is a set of conditional probability distributions P, each representing the distribution of one random variable X_i given all of its parents in the DAG. That is, $\mathcal{P} \equiv \{P(X_i|pa(X_i)), \forall X_i\}$, where

* Corresponding author. Tel.: +34 913944431; fax: +34 913944606.
E-mail address: pmain@mat.ucm.es (P. Main).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.04.020

$pa(X_i) \subset \{X_1,\ldots,X_{i-1}\}$. When X_i has no incoming arcs (no parents), $P(X_i|pa(X_i))$ stands for the marginal $P(X_i)$.

The joint probability distribution of a BN can be defined in terms of the elements of $\mathcal{P}$, as the product of the conditional and marginal probability distributions:

$$P(\mathbf{X}) = \prod_{i=1}^{n} P(X_i|pa(X_i)). \tag{1}$$

BNs have been studied by authors such as Jensen and Nielsen (2007), Lauritzen (1996) and Pearl (1988) among others.

It is common to consider BNs with discrete variables. Nevertheless, it is possible to work with some continuous distributions. For example, a GBN is a BN where all the variables of the model are Gaussian. Specifically, in a GBN the joint probability density of $\mathbf{X} = \{X_1,\ldots,X_n\}$ is a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$f(\mathbf{x}) = (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}. \tag{2}$$

Here $\boldsymbol{\mu}$ is the n-dimensional mean vector and $\boldsymbol{\Sigma}$ the $n \times n$ positive definite covariance matrix.

Alternatively, the joint density can be factorized using the conditional probability densities for every X_i $(i = 1,\ldots,n)$ given its parents. These are univariate normal distributions, and can be obtained from the joint density as

$$f(x_i|pa(x_i)) \sim N\left(x_i|\mu_i + \sum_{j=1}^{i-1}\beta_{ji}(x_j-\mu_j), v_i\right)$$

where μ_i is the mean of X_i, β_{ji} are the regression coefficients of X_i with respect to $X_j \in pa(X_i)$, and v_i is the conditional variance of X_i given its parents. Note that $\beta_{ji} = 0$ if and only if there is no link from X_j to X_i.

From the conditional specification it is also possible to get the parameters of the joint distribution. The means μ_i are the elements of the n-dimensional mean vector $\boldsymbol{\mu}$, and the covariance matrix $\boldsymbol{\Sigma}$ can be obtained with $\{v_i\}$ and $\{\beta_{ji}\}$ as follows. Let $\mathbf{D}$ be a diagonal matrix with the conditional variances $\mathbf{v}^T = (v_1,\ldots,v_n)$, $\mathbf{D} = diag(\mathbf{v})$. Let $\mathbf{B}$ be a strictly upper triangular matrix with the regression coefficients β_{ji}. Then the covariance matrix $\boldsymbol{\Sigma}$ can be computed as

$$\mathbf{\Sigma} = [(\mathbf{I}-\mathbf{B})^{-1}]^T\mathbf{D}(\mathbf{I}-\mathbf{B})^{-1} \tag{3}$$

For details, see Shachter and Kenley (1989).

As we remarked in Section 1, the conditional parameters, $\{v_i\}$ and $\{\beta_{ji}\}$ may be easier to specify by experts than $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. However, the two representations are completely equivalent when defining a GBN.

Now we introduce a working example of a GBN.

Example 1. Our sample problem concerns the amount of time that a machine will work before failing. The machine is made up of 7 elements, each with its own random time to failure X_i $(i = 1,\ldots,7)$. The elements are connected as shown in Fig. 1; the regression coefficients between variables are written next to the arcs.

The time that each element continues working is given by a normal distribution. The joint probability distribution of $\mathbf{X} = \{X_1,X_2,\ldots,X_7\}$ is a multivariate normal distribution.

Experts with the machine give the following parameters:

$$\boldsymbol{\mu} = \begin{pmatrix}1\\3\\2\\1\\4\\5\\8\end{pmatrix} \quad \mathbf{B} = \begin{pmatrix}0&0&0&1&0&0&0\\0&0&0&2&2&0&0\\0&0&0&0&1&0&0\\0&0&0&0&0&2&0\\0&0&0&0&0&2&0\\0&0&0&0&0&0&1\\0&0&0&0&0&0&0\end{pmatrix} \quad \mathbf{D} = \begin{pmatrix}1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&2&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&4&0&0\\0&0&0&0&0&1&0\\0&0&0&0&0&0&2\end{pmatrix}$$

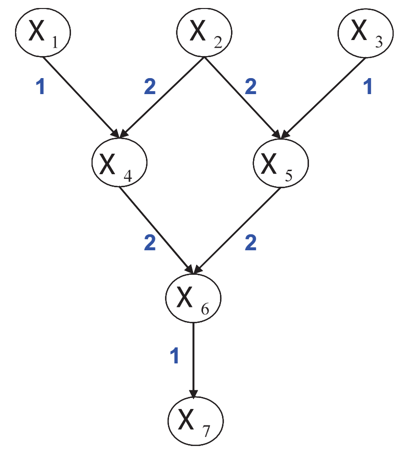

Fig. 1. DAG of the GBN in Example 1.

where $\mathbf{B}$ is the strictly upper triangular $(j < i)$ matrix of regression coefficients β_{ji} given in the DAG, and $\mathbf{D}$ is the diagonal matrix of conditional variances $\mathbf{v}^T = (v_1,\ldots,v_n)$, $\mathbf{D} = diag(\mathbf{v})$. Computing the joint parameters with Eq.(3) yields the random variable $\mathbf{X} \sim N(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix}1\\3\\2\\1\\4\\5\\8\end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix}1&0&0&1&0&2&2\\0&1&0&2&2&8&8\\0&0&2&0&2&4&4\\1&2&0&6&4&20&20\\0&2&2&4&10&28&28\\2&8&4&20&28&97&97\\2&8&4&20&28&97&99\end{pmatrix}$$

We now have both representations of the GBN to work with.

3. Sensitivity in GBNS

To build a BN is a difficult task, and expert knowledge is necessary to define the model. Furthermore, as discussed in the Introduction, even the parameters offered by experts may be inaccurate.

Sensitivity analysis is a general technique for evaluating the effects of inaccuracies in model parameters on the conclusions.

In a BN, the desired output is the marginal distribution of an interesting variable. This function is computed from the quantitative parameters that specify the BN. The result will be sensitive to inaccuracy in any parameter, but not all parameters require the same level of accuracy. Typically, some variables have more impact on the network's output than others (van der Gaag et al., 2007).

The model's sensitivity can be measured by varying one parameter while keeping the others fixed. This method is known as *one-way sensitivity analysis*. The output can also be studied while simultaneously changing a set of parameters, known as *n-way sensitivity analysis* (van der Gaag et al., 2007).

In recent years, a few authors have published sensitivity analyses of BNs (Bednarski, Cholewa, & Frid, 2004; Chan & Darwiche, 2005; Coupé, van der Gaag, & Habbema, 2000; Kjærulff & van der Gaag, 2000; Laskey, 1995). All of these works, while useful, deal exclusively with discrete BNs.

The sensitivity of GBNs has been studied only by Castillo and Kjærulff (2003) and Gómez-Villegas, Main, and Susi (2007, 2008). Castillo and Kjaerulff proposed a one-way sensitivity analysis (based on Laskey, 1995) investigating the impact of small changes in the network parameters.

Our first paper on this topic (Gómez-Villegas et al., 2007) is also a one-way sensitivity analysis. It differs from that of Castillo and Kjærulff (2003) in that we evaluated a global sensitivity measure