Information Sciences 275 (2014) 115–126

Contents lists available at ScienceDirect

Information Sciences

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

Sensitivity to evidence in Gaussian Bayesian networks using mutual information

Miguel Angel Gómez-Villegas [a,*], Paloma Main [a], Paola Viviani [b]

[a] Department of Statistics and O.R., Complutense University of Madrid, Spain
[b] Department of Public Health, Faculty of Medicine, Pontificia Universidad Católica de Chile, Chile

ARTICLE INFO

Article history:
Received 22 September 2011
Received in revised form 26 October 2013
Accepted 9 February 2014
Available online 19 February 2014

Keywords:
Entropy
Evidence propagation
Gaussian Bayesian network
Mutual information
Sensitivity analysis

ABSTRACT

We introduce a methodology for sensitivity analysis of evidence variables in Gaussian Bayesian networks. Knowledge of the posterior probability distribution of the target variable in a Bayesian network, given a set of evidence, is desirable. However, this evidence is not always determined; in fact, additional information might be requested to improve the solution in terms of reducing uncertainty. In this study we develop a procedure, based on Shannon entropy and information theory measures, that allows us to prioritize information according to its utility in yielding a better result. Some examples illustrate the concepts and methods introduced.

1. Introduction

A Bayesian network is a probabilistic graphical model that represents the conditional dependencies among a set of random variables through a directed acyclic graph (DAG). Bayesian networks have become an increasing popular representation for reasoning under uncertainty and are widely applied to diverse fields, such as medical diagnosis, image recognition, and decision-making systems, among many others.

Formally, a Bayesian network consists of qualitative and quantitative parts. The quantitative part is given by a DAG, whose nodes represent random variables that may be observable, latent, or a target variable of interest. The qualitative part, specifies the conditional probability distribution for each node given its parents; this allows us to compute the joint probability distribution of the model.

The aim of Bayesian network analysis is usually to obtain the conditional probability distribution of a target variable when a set of observable variables (evidence values) is available. Sometimes the variables defined as evidence are fixed in advance but other times they vary from model to model.

In this context, sensitivity analysis is a method for investigating the relationship between network inputs and the conditional distribution of the target variable, for which inputs can be the parameters considered in the conditional probability distribution or actual values taken by the observed variables. There is a large body of literature dealing with sensitivity analysis techniques for Bayesian networks. Most studies have addressed discrete Bayesian networks. For example, Malhas and Al Aghbari [16] introduced a score based on mutual information increases to discover new interesting patterns. Chan and Darwiche [4] presented a distance measure between the original distribution and a new one in which the parameters have

* Corresponding author. Tel.: +34 91 3944428; fax: +34 91 3944606.
E-mail addresses: ma_gv@mat.ucm.es (M.A. Gómez-Villegas), pmain@mat.ucm.es (P. Main), paolav@mat.puc.cl (P. Viviani).

http://dx.doi.org/10.1016/j.ins.2014.02.025
0020-0255/

been changed. Laskey [14] measured sensitivity by computing the partial derivatives of output probabilities with respect to given parameters. Kostal et al. [13] proposed measures of statistical dispersion based on Shannon and Fisher information. Castillo and Kjaerulff [3] developed a sensitivity analysis for Gaussian Bayesian networks (GBNs) using partial derivatives and symbolic propagation. Gómez-Villegas et al. [7–10] used Kullback Leibler divergence as a measure of sensitivity in GBNs.

Here we focus on a different aspect of sensitivity analysis. As mentioned previously, the set of evidence variables is not specified in advance in many real-life problems. In fact, it is usual practice to try to collect as much information as possible. However, this information always has an associated cost, so it may be desirable to evaluate which of all the available variables are most informative and useful for obtaining the best results. A very important assumption made in this paper is that a *better result* is achieved if the conditional probability distribution of the target variable has the lowest uncertainty, that is, the lowest entropy. Thus, we use information theory to provide tools to prioritize the available information to reduce the uncertainty of the target variable as far as possible.

The remainder of the article is structured as follows. In Section 2 we briefly review GBNs and show how propagation of observable values can be performed in this case. We also introduce our working example. Section 3 presents some general concepts of entropy, mutual information, and normalized measures. In Section 4, we first propose a procedure to study the sensitivity to evidence in GBNs and then perform a sensitivity analysis on our working example. The second contribution of the paper is presented in Section 5, which is an extension of the sensitivity analysis proposed above but incorporating normalized measures. Results are presented for the working example and a supplementary example. Finally, in Section 6 we draw conclusions.

2. Gaussian Bayesian networks

In this section we first recall the definition of a general Bayesian network and then the special case of a GBN. We also present the methodology for evidence propagation in GBNs.

2.1. Definition: Bayesian network

A *Bayesian network* is a pair $(\mathcal{G}, \mathcal{P})$, where $\mathcal{G}$ is a directed acyclic graph (DAG) with one node for each random variable of $\mathbf{X} = \{X_1, \ldots, X_n\}$ and edges that represent probabilistic dependencies between them.

$\mathcal{P} = \{p(x_1|pa(x_1)), \ldots, p(x_n|pa(x_n))\}$ is a set of conditional probability distributions and $pa(x_i)$ is the set of parents of node X_i in $\mathcal{G}$. From $\mathcal{P}$, the associated joint probability distribution for $\mathbf{X}$ is defined as

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

The type of random variables, X_i, considered in the problem defines whether we are dealing with discrete, Gaussian, or mixed Bayesian networks. In this paper we develop results for GBNs that are based on continuous variables; these have been studied by Castillo et al. [2], Cowell et al. [6] and Gómez-Villegas et al. [7], among others.

2.2. Definition: GBN

GBNs are a subclass of Bayesian networks in which the joint probability density of $\mathbf{X}$ is a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, that is,

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

where $\boldsymbol{\mu}$ is the n-dimensional mean vector, $\boldsymbol{\Sigma}$ is the positive definite $n \times n$ covariance matrix, $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$, and $\boldsymbol{\mu}^T$ the transpose of $\boldsymbol{\mu}$.

According to the normal distribution properties and the factorization presented in (1), in a GBN the joint probability density can be specified also as a product of conditional probability densities, each of which corresponds to a univariate normal distribution.

2.3. Evidence propagation in a GBN

In real-life problems, information about the state of one or more variables of a Bayesian network, known as evidence variables, may be available. If so, probability distributions for the rest of the variables in the network can be updated given the observed values. This process is called *evidence propagation*.

Different algorithms have been proposed for evidence propagation in GBNs. Here, we consider an incremental method developed by Castillo et al. [2]. This consists of computing the conditional probability density of a normal distribution after introducing one evidential variable at a time. We consider the set of non-evidential variables $\mathbf{Y}$ and the evidential variables $\mathbf{E}$. Then $\mathbf{X}$ can be written as the partition $\mathbf{X} = (\mathbf{Y}, \mathbf{E})$, and the conditional distribution of $\mathbf{Y}$ given $\mathbf{E} = \mathbf{e}$ is a multivariate normal distribution with parameters