Available online at www.sciencedirect.com
ScienceDirect
International Journal of Approximate Reasoning
48 (2008) 659–666
INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

Bayesian networks with a logistic regression model for the conditional probabilities

Frank Rijmen *

Department of Clinical Epidemiology and Biostatistics, VU Medical Centre, Amsterdam, The Netherlands

Received 26 March 2007; received in revised form 9 January 2008; accepted 9 January 2008
Available online 20 January 2008

Abstract

Logistic regression techniques can be used to restrict the conditional probabilities of a Bayesian network for discrete variables. More specifically, each variable of the network can be modeled through a logistic regression model, in which the parents of the variable define the covariates. When all main effects and interactions between the parent variables are incorporated as covariates, the conditional probabilities are estimated without restrictions, as in a traditional Bayesian network. By incorporating interaction terms up to a specific order only, the number of parameters can be drastically reduced. Furthermore, ordered logistic regression can be used when the categories of a variable are ordered, resulting in even more parsimonious models. Parameters are estimated by a modified junction tree algorithm. The approach is illustrated with the Alarm network.

Keywords: Bayesian networks; Logistic regression; Generalized linear models; Restricted conditional probabilities

0. Introduction

In a probabilistic graphical model, random variables are represented by nodes, and the (absence of) edges between nodes represent conditional (in)dependence relations. Apart from offering an appealing way to represent models visually, efficient computational schemes can be constructed by working on the graph associated with a probabilistic model [7].

Recently, research has been focused on structural learning. That is, how can we identify a set of conditional dependence relations that is both parsimonious and provides an adequate fit to a given dataset? Several procedures have been proposed in the literature (for reviews, see [2,8,12]). In this paper on the other hand, we focus on learning the parameters of an inferred (or a priori given) network structure. We consider Bayesian networks for discrete variables, where dependence relations are encoded through directed edges. More specifically, we show how the number of effective parameters of the network can be reduced by adopting a logistic regression framework for modelling the conditional dependence relations.

* Address: Rosedale Road, Princeton, NJ 08541, United States.
E-mail address: frijmen@ets.org

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2008.01.001

In a Bayesian network, the probability distribution of a set of random variables $\mathbf{X} = (X_1, \ldots, X_M)'$ can be recursively factorized as

$$\Pr(\mathbf{X}) = \prod_{m=1}^{M} \Pr(X_m|\text{pa}(X_m)), \tag{1}$$

where $\text{pa}(X_m)$ is the set of random variables that are parents of X_m in the directed acyclic graph that is associated with $\Pr(\mathbf{X})$. Learning the parameters of a Bayesian network for discrete variables hence comes down to learning the parameters that govern the conditional probability tables $\Pr(X_m|\text{pa}(X_m))$. Usually, these conditional probability tables are not restricted beyond the obvious restriction that $\sum_{j=1}^{J_m} \Pr(X_m = j|\text{pa}(X_m)) = 1$, where J_m is the number of distinct values X_m can take. In some model families, equality restrictions between conditional probability tables are encountered as well. For example, in hidden Markov type of models, a default assumption is that the conditional probability tables do not change over time. Regardless of the latter type of restrictions, each additional parent adds a dimension to the conditional probability table, so that the number of parameters increases exponentially with the number of parents when these conditional probabilities are not further restricted. Consequently, for small to moderately sized data sets, parameters can only be reliably estimated for fairly simple network structures. In the Bayesian networks field, this problem is most often tackled by incorporating "prior information," leading to either penalized maximum likelihood estimation or a fully Bayesian approach. When prior information is available through substantive knowledge or previous studies (rather than the prior 'knowledge' that extreme probabilities are unlikely), this is quite a reasonable approach.

In this paper, an alternative approach to tackle the estimation problem is proposed. More specifically, the number of parameters is controlled by modelling the conditional probabilities as a function of a limited set of parameters using logistic regression.

1. Modelling the conditional probabilities with multinomial logistic regression

Let y_i, $i = 1, \ldots, n$ denote a set of independent realizations of a categorical outcome variable Y, and $\mathbf{z}_i$ the corresponding vector of realizations of p covariates. Then, a multinomial logistic regression model can be specified as follows (e.g. [4]):

– y_i is a realization from a multinomial distribution

$$\Pr(Y_i = j) = \pi_{ij} \quad \text{with} \quad \sum_j \pi_{ij} = 1 \tag{2}$$

– The parameter vector $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iJ-1})'$ (π_{iJ} is redundant since $\sum_j \pi_{ij} = 1$) is related to the *linear predictor* $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iJ-1})'$ via the multinomial link function:

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \eta_{ij}. \tag{3}$$

– $\boldsymbol{\eta}_i = \mathbf{Z}_i\boldsymbol{\beta}$, where $\mathbf{Z}_i$ is the so-called design matrix of size $J - 1$ by p constructed from $\mathbf{z}_i$; and $\boldsymbol{\beta}$ is a p-dimensional parameter vector.

The multinomial logistic regression model can be integrated into a Bayesian network by modelling each conditional probability table $\Pr(X_m|\text{pa}(X_m)$ of a particular Bayesian network with a multinomial logistic regression model, where X_m is the outcome variable and the design matrix $\mathbf{Z}_{mi}$ is constructed from $pa(X_m)$.

A Bayesian network without restrictions on the conditional probability tables is obtained by constructing $\mathbf{Z}_{mi}$ from $\text{pa}(X_m)$ as follows. For each possible configuration s on $\text{pa}(X_m), s = 1, \ldots, S = \prod_{k:X_k \in \text{pa}(X_m)} J_k$, a dummy variable is defined. For each case i, the covariate vector $\mathbf{z}_{im} = (z_{im1}, \ldots, z_{imS})'$ is defined as an indicator vector with $z_{ims} = 1$ if configuration s is observed, and $z_{ims} = 0$ otherwise. The $(J_m - 1)$ by $(J_m - 1) \times S$ design matrix $\mathbf{Z}_{im}$ is constructed from $\mathbf{z}_{im}$ as