Decision Support Systems 87 (2016) 69–79

Contents lists available at ScienceDirect

Decision Support Systems

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

An empirical study of Bayesian network parameter learning with monotonic influence constraints

Yun Zhou[a,b,*], Norman Fenton[b], Cheng Zhu[a]

[a]Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, China
[b]Risk and Information Management (RIM) Research Group, Queen Mary University of London, United Kingdom

ARTICLE INFO

Article history:
Received 16 December 2015
Received in revised form 9 May 2016
Accepted 9 May 2016
Available online 20 May 2016

Keywords:
BN parameter learning
Monotonic influences
Exterior constraints
Experiments on publicly available BNs
Real medical study

ABSTRACT

Learning the conditional probability table (CPT) parameters of Bayesian networks (BNs) is a key challenge in real-world decision support applications, especially when there are limited data available. A conventional way to address this challenge is to introduce domain knowledge/expert judgments that are encoded as qualitative parameter constraints. In this paper we focus on a class of constraints which is naturally encoded in the edges of BNs with monotonic influences. Experimental results indicate that such monotonic influence constraints are widespread in practical BNs (all BNs used in the study contain such monotonic influences). To exploit expert knowledge about such constraints we have developed an improved constrained optimization algorithm, which achieves good parameter learning performance using these constraints, especially when data are limited. Specifically, this algorithm outperforms the previous state-of-the-art and is also robust to errors in labelling the monotonic influences. The method is applied to a real world medical decision support BN where we had access to expert-provided constraints and real hospital data. The results suggest that incorporating expert judgments about monotonic influence constraints can lead to more accurate BNs for decision support and risk analysis.

1. Introduction

Bayesian networks (BNs) have become increasingly popular in the AI field during the last two decades because of their ability to model probabilistic dependent relationships among variables in many real-world problems. A BN model consists of two components: a network structure and a set of conditional probability tables (CPTs) whose entries are considered as parameters.

In real-world decision support problems that we wish to model as BNs, there are typically limited or no relevant data. In such situations attempts to learn BN structures purely from data are unlikely to result in useful models. For example, even 500 data points (which in many real-world situations is a very large sample) is nowhere near enough to learn the structure of a very small BN such as the well-known Asia BN that has just 8 nodes and 8 edges in total. Using the pure data-based structure learning algorithm [18] in this example results in more than half of the learnt edges being different from the ground truth. The scarce data problem is typical of many real-world decision support problems in which we have nevertheless used BN models effectively, by exploiting expert domain knowledge. Specifically, the decision support problems addressed include:

- determine whether or not to provide a specific type of intervention for a given psychiatric patient [10].
- determine whether or not a limb should be amputated given a patient's specific pathology [53].
- determine whether a given prisoner with a background of violence can be safely released into the community [13].
- determine which of two alternative medical tests optimises the balance between accuracy, safety and cost [21].
- determine how and when to place bets on football matches to 'beat the bookmakers' [11].

In all of these problems, limited data were available (both in terms of size of data and complete absence of data for some key variables), but we had access to relevant domain experts who were able to provide the BN structure (including causal relationships involving unobserved variables) and insights into the conditional probability table (CPT) parameters where there was little or no data. However, although there has been some progress in attempts

* Corresponding author.
E-mail addresses: zhouy.nudt@gmail.com (Y. Zhou), n.fenton@qmul.ac.uk (N. Fenton), zhucheng@nudt.edu.cn (C. Zhu).

http://dx.doi.org/10.1016/j.dss.2016.05.001
0167-9236/

to make more systematic the methods helping experts define BN structures (see, e.g. Refs. [10] and [24] the process for fully defining the CPTs, i.e. eliciting the BN parameters) from experts has been largely ad hoc. The objective of this paper is to demonstrate a more systematic and rigorous approach to combining expert knowledge and data to achieve more accurate parameter elicitation in such models. Hence, to clarify the scope, the paper is focused on the following common scenario:

> A BN structure has been hand-crafted by domain experts to model a real-world decision support problem. A small amount of data relevant to the model is available. The challenge is to build the model parameters by combining the limited data with domain knowledge about the parameters.

In addition to the examples described above an increasing number of decision support problems (medical, financial and safety) fit with this scenario [22], and so there is a genuine demand for improved solutions. It is also important to note that, because we are restricting our discussion to BNs whose structures have been hand-crafted by experts, the scope is limited to relatively 'small' BNs (generally expert defined BNs with fewer than 100 nodes), although our experiments do include some larger BNs.

The simplest parameter learning approach is maximum likelihood estimation (MLE). However, this method usually fails to find good estimates for parameters with few data points (in some complex BNs, there is an explosion of variable state configurations, we might not have enough training data in some specific variable state configurations even in cases where big-data is available). To address this researchers developed the maximum a posteriori probability (MAP) approach by introducing a Dirichlet parameter prior, which we discuss in Section 2. However, as we also discuss in Section 2, experts tend to feel more comfortable providing qualitative or semi-numerical judgments with less cognitive effort. Such judgments are expressed as constraints between parameters of interest, and are more easily elicited from experts than corresponding point-wise estimates.

In this paper, we focus on an important class of such constraints elicited from monotonic influences (also known as qualitative influences [48] or qualitative monotonicities [1]), which are naturally encoded in the edges/structures of BNs. A monotonic influence is one where the increase (or decrease) of one variable will monotonically change the value of another variable. This kind of influence can be directly elicited from the BN structures, and can be easily converted into associated parameter constraints, which we refer to as monotonic influence constraints. These constraints are exterior parameter constraints (relations between parameters from different conditional distributions). For a simple example, in "Smoke → Cancer", it is widely accepted that people who smoke have a higher risk of getting cancer than those who do not. Thus:

$$P(Cancer = true|Smoke = true) \geq P(Cancer = true|Smoke = false)$$

is an example of a monotonic influence constraint.

When the training data is limited, incorporating such exterior constraints from experts could help the BN parameter learning. In this paper, we investigate the extent to which such monotonic influences and their generated exterior constraints are present in a set of real-world BNs, and provide a simple improved constrained optimization algorithm for parameter estimation with these constraints.

The paper is organized as follows. In Section 2, we discuss related work in BN parameter learning with limited data. In Section 3, we introduce the BN parameter learning notation to be used throughout this paper. In Section 4, we describe the monotonic influences and the improved parameter learning method. In Section 5, we report on the experiments of 12 different real-world BNs. In Section 6, we present the results of applying the method to a real world medical decision support BN. Our conclusions are in Section 7.

2. Related works

There are several methods for handling parameterization with limited or no relevant data, described in a rich literature of books, articles and software packages, which are briefly summarized in Refs. [15,22,37,41]. Of these, expert knowledge/judgments are widely used in real-world BN construction [3,6,11], especially in medical decision support applications [12,25,30,34,52,53].

However, expert elicitation is expensive, time-consuming and sometimes error-prone [38], because the number of parameters increase exponentially with the number of nodes in the BN. Therefore, the challenge has mainly been addressed using methods that minimize the number of elicited parameters. The Noisy-OR [14] and Noisy-MAX [43] are examples of methods to reduce the number of elicited parameters, based on the independence of causal influences (ICI) assumption [54]. Extensions of these models include the Ranked Node [23] and NIN-AND tree [49,50] models.

To address the problem that some parameters have zero observations in limited training data, a Dirichlet parameter prior is introduced for them. Experts are required to provide Dirichlet hyperparameters. In the BDeu prior, experts are only needed to provide the equivalent sample size parameter [28]. Guidance in choosing the value of equivalent sample size is well studied [47]. However, elicited hyperparameters of ICI models and Dirichlet distributions are both numerical, which means they are quantitative knowledge. Previous work has shown that eliciting qualitative or semi-numerical judgments is easier than collecting numerical values of CPTs [29]. Parameter constraint [16] is an important class of such qualitative judgments. For example, the statement "the probability of people getting cancer is very low" is such a parameter constraint.

Several models have been proposed to integrate parameter constraints and improve the learning accuracy. The most popular is the constrained convex optimization (CO) formulation [5,6,7,33,39]. These algorithms seek the global optimal estimation (maximal log likelihood) with respect to the parameter constraints. The parameters also can be estimated by the Monte Carlo method [9], where only the samples that consist of the constraints are kept. Recently, auxiliary BN models [55,56,58] have been developed for solving this problem. In this approach, the target parameters, data observations and elicited constraints are all modelled as nodes in the auxiliary BNs. Thus, the parameters are estimated via the inference in the auxiliary BNs. However, constraints discussed in these models are not elicited from qualitative monotonic influences, and are usually expensive to elicit.

An alternative approach to reducing the burden of expert elicitation is to find monotonic influences in some edges of BNs, and use them to generate exterior parameter constraints. BNs that are fully specified by monotonic influences are referred to as Qualitative Probabilistic Networks (QPNs) [48]. An efficient sign-propagation algorithm is achieved by restricting the maximal number of node-sign changes during the inference [17,45]. The inference results answer the question of how observations of some variables change the probability distributions of other variables. The combination of QPNs and BNs is referred to as Semi-Qualitative Probabilistic Networks (SQPNs) [44], which means parts of the variables are represented by joint probability tables rather than qualitative influences. Inference and learning in SQPNs is discussed in later work [4].

As in previous work [1,19,20,26], in this paper, we only use signs of qualitative probabilistic networks and their generated monotonic influence constraints to constrain the probabilities in the standard BN parameter learning. Thus, experts are only required to identify which edges in the BN have such qualitative monotonicity property.