Expert Systems With Applications 49 (2016) 1–19

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Inference in hybrid Bayesian networks with large discrete and continuous domains

Junichi Mori[a], Vladimir Mahalec[b,*]

[a] Nippon Steel & Sumitomo Metal Corporation, Kimitsu Works, Department of Computer System, Kimitsu City, Chiba Prefecture 299-1141, Japan
[b] McMaster University, Department of Chemical Engineering, Hamilton, Ontario L8S 4L7, Canada

ARTICLE INFO

Keywords:
Large domain of discrete and continuous variables
Decision tree
Hybrid Bayesian network
Belief propagation algorithm
Context-specific independence

ABSTRACT

Inference in Bayesian networks with large domain of discrete variables requires significant computational effort. In order to reduce the computational effort, current approaches often assume that discrete variables have some bounded number of values or are represented at an appropriate size of clusters. In this paper, we introduce decision-tree structured conditional probability representations that can efficiently handle a large domain of discrete and continuous variables. These representations can partition the large number of values into some reasonable number of clusters and lead to more robust parameter estimation. Very rapid computation and ability to treat both discrete and continuous variables are accomplished via modified belief propagation algorithm. Being able to compute various types of reasoning from a single Bayesian network eliminates development and maintenance issues associated with the use of distinct models for different types of reasoning. Application to real-world steel production process data is presented.

1. Introduction

This work has been motivated by the need to optimize production of steel plates manufacturing. Steel plates manufacturing is a complex, multi-stage process. A manufacturing plant often produces several thousands of different steel plate SKUs. Each stage of the manufacturing is not a fully deterministic process. Sometimes there may be defects at some stage, which are then corrected by modifying the manufacturing process. In other instances, a customer may order a new steel plate, something that has not been manufactured yet. Due to these circumstances, the exact times required to produce any specific SKU is not known. Production planning and scheduling models require that we estimate the production times for each grade of steel plates. Such estimate can be made from a Bayesian network representing the manufacturing plant and probabilities of processing a steel plate at each stage of the manufacturing. Due to the complexity of the steel manufacturing plant, it is not possible to use a first-principle model of the plant to construct such Bayesian network. In this paper we use Bayesian statistics to construct the most likely Bayesian network for such complex manufacturing process. Having constructed Bayesian network, we then proceed to estimate most likely production times for each grade of steel plates. The challenging nature of the problem is magnified by the large number of different grades of steel plates. This paper presents new inference algorithm in Bayesian network with large domain discrete variables to enable us to:

1. Estimate the probability distributions of production time from historical data with large domain discrete variables and continuous variables.
2. Deal with unobservable (unavailable) variables such that we have a single model and avoid multiple models that meet with specific problems.

We present a new method for constructing most likely structure of the Bayesian network representing a complex manufacturing process, e.g. manufacturing of steel plates. Such networks contain a large number of discrete variables and also contain continuous variables parent nodes which have discrete variables children nodes. An application to a steel plate manufacturing process, which produces a large number of distinct steel plates and also has uncertain production times at different manufacturing steps, demonstrates that the proposed method can successfully compute inferences for very large hybrid Bayesian networks.

In order to learn the tree structured CPTs, scores such as Bayesian scores are often used as objective functions. However, greedy hill climbing approach cannot be used since it can easily get stuck in a local minimum at the early stage. Some kinds of approach to avoid local minimum as much as possible are proposed, but they are computationally expensive. Therefore, we employ decision trees algorithm

* Corresponding author. Tel.: +1 905 525 9140 ext.26386.
E-mail addresses: mori.nn4.junichi@jp.nssmc.com (J. Mori), mahalec@mcmaster.ca (V. Mahalec).

http://dx.doi.org/10.1016/j.eswa.2015.11.019
0957-4174/

(Breiman, Friedman, Olshen, & Stone, 1984) based on classification trees in order to learn the tree structured CPTs. Classification trees predict the dependent variables following decisions in the tree from the root node down to the leaf node. Since the classification trees group the values to capture important distinctions of continuous or discrete variables, this method can be used to construct the context-specific CPTs in the hybrid Bayesian networks. The classification tree classifies discrete variables into a small number of subsets so that the values of continuous or discrete child nodes can be distinguished well. If Bayesian networks include continuous parent nodes with discrete child nodes, the corresponding continuous variables can be discretized as finely as needed, because the domain size of discretized variable does not increase the number of parameters in intermediate factors due to decision-tree structured CPTs. Since the classification algorithms are typically greedy ones, the computational cost is relatively small. Consequently, the intermediate factors can be described compactly using a simple parametric representation called the canonical form.

We also introduce the decision-tree structured CPT based inference algorithm in Bayesian network, which employs belief propagation algorithm to deal with hybrid networks with large domain discrete variables. In order for multiplying and marginalizing factors during belief propagation, novel types of operations to dynamically construct CPTs are introduced. In order to carry out other types of inferences, such as causal, diagnostic, intercausal and mixed reasoning, we employ the loopy belief propagation in the decision-tree structured CPT based Bayesian networks.

The organization of the article is as follows. Review of the related prior work is presented in Section 2. Section 3 describes construction of decision-tree structured CPTs for hybrid Bayesian networks with large domain discrete variables. Section 4 describes the detailed inference algorithm including operations of product and marginalization of factors. The presented method is applied to the steel production processes data in Section 5. Finally, the conclusions are presented in Section 6.

2. Review of prior related works

In this section we review prior related works and discuss the limitations with respect the size of the problem and complexity of computation.

Let us first review prior related works and discuss the capabilities of the proposed methods with respect the size of the problem and complexity of computation.

Probabilistic graphical models are popular for representing conditional independencies among random variables under system uncertainty. Such models are comprised of nodes representing random variables and the links between the nodes which express probabilistic relationships among the corresponding random variables. Two major classes of graphical models are *Bayesian networks* and *Markov random fields*. Bayesian networks are also called *directed graphical models* since the links of the graphs represent direct dependence among the variables and are described by arrows between links. Markov random fields are also called *undirected graphical models* since they provide a simple definition of independence among random variables and do not have a particular directionality indicated by arrows (Bishop, 2006; Pearl, 1988). Both graphical models are popular in the machine learning community and have been applied to various fields including medical diagnostics, speech recognition, gene modeling, cancer classification, target tracking, sensor validation, and reliability analysis.

In particular, Bayesian network has been widely used for systems including many uncertainties. For instance, scenario analysis under changing conditions is implemented by means of Bayesian inference techniques, since Bayesian network performs well in uncertainty environment (Buyukozkan, Kayakutlu, & Karakadlar, 2015; Cai, Sun, Si, & Yannou, 2011). Bayesian network is also applied to predicting the risk of software development or maintenance projects, because it is suitable for representing the knowledge of experts under uncertainty of conditions (Melo & Sanchez, 2008; Perkusich & Soares, 2014). As these applications indicate, Bayesian network is a powerful tool for knowledge representation and reasoning under uncertainties since it can visually represents the probabilistic relationships among measured and unmeasured variables.

Each node in a Bayesian network is associated with conditional probability distributions (CPD). The most common representations of CPDs are conditional probability tables (CPTs), which specify marginal probability distributions for each combination of values of its discrete parent nodes. The number of parameters required to represent CPTs grows exponentially both with the number of discrete variables and with the cardinality of discrete variables. In order to reduce the number of parameters, context-specific independence representations have been proposed (Boutilier, Friedman, Goldszmidt, & Koller, 1996). Furthermore, an efficient inference algorithm that exploits context-specified independence has also been proposed (Poole & Zhang, 2003). As for identification of parameters of context-specific independence, learning methods such as tree-structured CPTs (Friedman & Goldszmidt, 1996) and graph-structured CPTs (Chickering, Heckerman, & Meek, 1997) have been developed. However, since learning structured CPTs is NP-hard problems, all of these methods assume that all discrete variables have a bounded number of values or that they are already grouped at an appropriate level of domain size. In the real world problem, discrete variables often have large domains and the task of grouping discrete values requires expert knowledge that enables us to identify a reasonable set of groups that well distinguish the values of discrete variables. In order to group the discrete values in a Bayesian network learning, *attribute − value hierarchies* (AVHs) which capture meaningful groupings of values in a particular domain are integrated with the tree-structured CPTs (DesJardins & Rathod, 2008). However, if large domain discrete variables do not contain hierarchal structures, AVHs cannot capture the useful abstracts of values in that domain. In addition, this model cannot handle the continuous variables without discretizing them. The authors, DesJardins and Rathod (2008) also do not apply AVH-derived CPTs to inference in Bayesian networks.

Sharma and Poole (2003) have proposed an inference method for Bayesian Networks containing CPTs which are represented as decision trees. The inference algorithm is based on variable elimination (VE) algorithm; the authors introduced operations for decision trees computations, namely multiplying factors and summing out variable from a factor. However, because the computational complexity of the exact inference algorithm such as VE grows exponentially with the size of the network, this method may not be appropriate for Bayesian networks in real world. In addition, computational cost of reconstructing decision trees as required to compute multiplication and marginalization of factors makes this method too expensive to apply to large decision trees. An alternative approach is to employ algebraic decision diagrams (ADDs) for the purpose of inference in Bayesian network with large domain discrete variables. For instance, ADDs have been used to represent factors and their multiplying and summing-out operations have been proposed (Chavira & Darwiche, 2007). In addition, structured message passing has been proposed to utilize powerful approximate inference algorithms such as cluster-graph Belief propagation (Gogate & Domingos, 2013). In the worst case, ADDs have the same space complexity as CPTs. To make matters worse, the factor operations of multiplication and summing-out are polynomial in time. It should be noted that all the above methods have not considered an application of decision-tree structured CPTs to hybrid Bayesian networks, where both discrete and continuous variables appear simultaneously.