International Journal of Approximate Reasoning 50 (2009) 472–484

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

Triangulation of Bayesian networks with recursive estimation of distribution algorithms

Txomin Romero [a,b,*], Pedro Larrañaga [c]

[a] Department of Computer Science and Artificial Intelligence, University of the Basque Country, Spain
[b] Donostia International Physics Center, Donostia – San Sebastian, Paseo Manuel de Lardizabal 4, 20018 Guipuzcoa, Spain
[c] Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Spain

ARTICLE INFO

Article history:
Received 5 September 2007
Received in revised form 18 September 2008
Accepted 18 September 2008
Available online 7 October 2008

Keywords:
Bayesian networks
Estimation of distribution algorithms
Triangulation
Inference
Junction tree

ABSTRACT

Bayesian networks can be used as a model to make inferences in domains with intrinsic uncertainty, that is, to determine the probability distribution of a set of variables given the instantiation of another set. The inference is an NP-hard problem. There are several algorithms to make exact and approximate inference. One of the most popular, and that is also an exact method, is the evidence propagation algorithm of Lauritzen and Spiegelhalter [S.L. Lauritzen, D.J. Spiegelhalter, Local computations with probabilities on graphical structures and their application on expert systems, Journal of the Royal Statistical Society B 50 (2) (1988) 157–224], improved later by Jensen et al. [F.V. Jensen, S.L. Lauritzen, K.G. Olesen, Bayesian updating in causal probabilistic networks by local computations, In Computational Statistics Quaterly 4 (1990) 269–282]. This algorithm needs an ordering of the variables in order to make the triangulation of the moral graph associated with the original Bayesian network structure. The effectiveness of the inference depends on the variable ordering. In this paper, we will use a new paradigm for evolutionary computation, the estimation of distribution algorithms (EDAs), to get the optimal ordering of the variables to obtain the most efficient triangulation. We will also present a new type of evolutionary algorithm, the recursive EDAs (REDAs). We will prove that REDAs improve the behaviour of EDAs in this particular problem, and that their results are competitive with other triangulation techniques.

1. Introduction

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be an n-dimensional random variable, where x_i is an instantiation of X_i. A PGM or *Probabilistic Graphical Model* $(S, \boldsymbol{\theta}_\mathbf{s})$ is a graphical structure $S = (\mathbf{X}, \mathscr{A})$ and a set of local parameters $\boldsymbol{\theta}_\mathbf{s}$. $\mathbf{X}$, a set of nodes, represents the system variables and $\mathscr{A}$, a set of arcs, the conditional dependence/independence relationships among the variables of the structure. A *Bayesian network* (BN) is a PGM where the graphical structure is a directed acyclic graph (DAG), X_i are random discrete variables (called *nodes*) and the set of parameters $\boldsymbol{\theta}_\mathbf{s} = (\theta_{ijk})$, where k goes from 1 to r_i, j from 1 to q_i and i from 1 to n, represents the local probability distributions over $\mathbf{X}$, i.e., θ_{ijk} is the conditional probability of X_i being in its kth value given that the set of its parents variables is in its jth-value. Finally, r_i is the number of different values of ith variable and q_i represents the number of possible instantiations for the set of parents of ith variable.

---

[*] Corresponding author. Address: Donostia International Physics Center, Donostia – San Sebastian, Paseo Manuel de Lardizabal 4, 20018 Guipuzcoa, Spain. Tel.: +34 943 018347; fax: +34 943 015600.

E-mail addresses: txomin.romero@ehu.es (T. Romero), pedro.larranaga@fi.upm.es (P. Larrañaga).

0888-613X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2008.09.002

The BN paradigm can be used as a model to make inferences in domains with intrinsic uncertainty. The factorisation of the joint distribution that a BN represents allows an efficient reasoning inside the model. Introductions and classic textbooks about BN include [10,18] and [30]. As a model to make inferences, we will try to obtain their best triangulation, but we will also use them in the estimation of distribution algorithms themselves.

The most direct way to make inference in a BN is to compute the marginalization over the not instantiated variables. The number of terms involved in the marginalization grows exponentially with the number of variables. Lauritzen and Spiegelhalter show [26] that, instead of calculating each joint probability separately in order to add them in a posterior step, we can group the common factors of all of them in order to make the calculation more efficient. For example, for the Asia network (Fig. 1), we have:

$$\begin{aligned} P(A,X,D) &= \sum_{T,E,L,B,S} P(A,T,X,E,D,L,B,S) \\ &= P(A)\sum_{T} P(T|A)\left[\sum_{E} P(X|E)\left[\sum_{L} P(E|T,L)\left[\sum_{B} P(D|E,B)\left[\sum_{S} P(L|S)P(B|S)P(S)\right]\right]\right]\right] \end{aligned} \tag{1}$$

We can rewrite Eq. (1) as

$$P(A,X,D) = \textstyle\sum_{T,E,L,B,S} \psi(A)\psi(T,A)\psi(X,E)\psi(E,T,L)\psi(D,E,B)\psi(L,S)\psi(B,S)\psi(S) \tag{2}$$

Here, each factor is not a conditional probability, but a *potential function*. The potential functions are, at first, the conditional probabilities, and their parameters are the variables that are connected in the graph. An useful procedure to discover which variables are related to each potential function is to moralize the graph, that is, to add an arc between the parent nodes of each node (see Fig. 2). Nevertheless, when we sum over S in expression $\psi(L,S)\psi(B,S)\psi(S)$ for all the values of S, we obtain a new factor (a new auxiliary potential function) depending on L and B, and between these nodes there is no arc. This is a problem due to the prior definition of potential function. We can solve this problem with the *triangulation* of the graph (see [26] for a more extensive explanation). A graph is triangulated if it has no cycles with a length greater than three without a cord. In the Asia network, it is enough to add $L-B$ arc in order to triangulate it (see Fig. 3), but in more complex networks it is not so easy. A good triangulation could make it possible to obtain a solution to some problems related to graphs in polynomial time instead of exponential time [3].

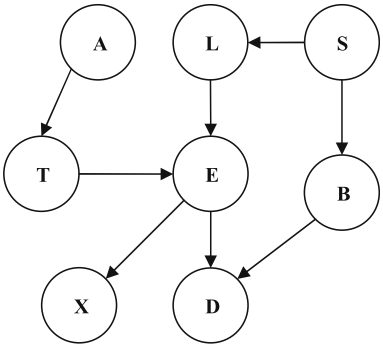

Fig. 1. The Asia Bayesian network.

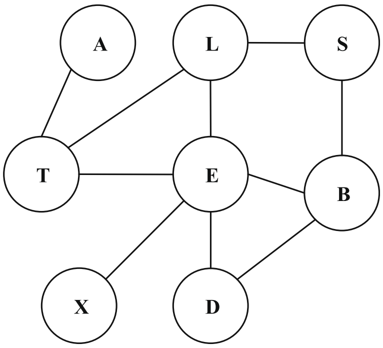

Fig. 2. The moralized Asia network.