

On-demand Thrifty Propagation for Belief Updating in Bayesian Networks-A Preliminary Report

Dan Wu

School of Computer Science
University of Windsor
Windsor Ontario
Canada N9B 3P4

Abstract

In this paper, we suggest a prototype of an on-demand thrifty global propagation scheme especially designed for belief updating in large and complex Bayesian networks.

1. Introduction

The Bayesian network (BN) model is a probabilistic graph model that has been successfully developed and applied in various domains for uncertainty management (Pearl 1988). It abstracts a problem domain using a set U of random variables and uses a *directed acyclic graph* (DAG) to encode the conditional independency information among U . One of the most important services that a BN can provide is belief updating which simply means calculating the posterior probability $p(x|E = e)$ for a variable $x \in U$ given that variables in $E \subset U$ are taking specific value e . The renowned *global propagation* (GP) method (Lauritzen & Spiegelhalter 1988; Jensen, Lauritzen, & Olesen 1990) is a successful approach for belief updating. Although the GP method performs quite well on many BNs in practice with less than or around 1000 nodes, it is questionable if it will be still effective when applied to large and complex domains.

In this paper, we suggest an on-demand thrifty approach for belief updating in large BNs which is based on the GP method but with novel features specially designed for large BNs. The computation needed for computing $p(x|E = e)$ is *on-demand* and *thrifty* in the sense that only the absolutely necessary minimal computation is carried out to obtain $p(x|E = e)$.

2. Global Propagation and Belief Updating

A *Bayesian network* defined over a set U of variables, written as $(\mathcal{D}, \mathcal{P})$, is a probabilistic graphical model where \mathcal{D} denotes a DAG and \mathcal{P} denotes a set of *conditional probabilistic distributions* (CPDs) such that each node x in \mathcal{D} (denoted $x \in \mathcal{D}$) corresponds one-to-one to a variable in U and each node is also associated one-to-one with a CPD $p(x|\pi_x)$ in \mathcal{P} , where π_x denotes the parents of x in \mathcal{D} . The product of the CPDs in \mathcal{P} defines a *joint probability distribution* (JPD) over U as $p(U) = \prod_{x \in \mathcal{D}} p(x|\pi_x)$ (Pearl 1988).

Copyright © 2005, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

The standard method for performing belief updating is the so-called *global propagation* method. Since it is well understood by the community, we refer the readers to (Lauritzen & Spiegelhalter 1988) for a complete exposition. The GP method is performed on a junction tree transformed from the DAG of a BN. The notion of evidence is to indicate that some variables in U are taking specific values from their respective domains. We use \mathbf{E} to denote this evidence and $v(\mathbf{E})$ to denote variables occurring in \mathbf{E} . Two pieces of evidence \mathbf{E} and \mathbf{E}' are *contradicting* if there exists a variable x such that $x \in v(\mathbf{E})$ and $x \in v(\mathbf{E}')$ but x are taking different values in \mathbf{E} and \mathbf{E}' . Otherwise, \mathbf{E} and \mathbf{E}' are *compatible*.

To summarize, applying the GP method to a junction tree *with no evidence observed* results in marginal distributions for cliques (of the junction tree) computed; applying the GP method to a junction tree *with evidence observed and incorporated* results in updated marginal distributions computed for cliques. The application of the GP method is *full scale* in the sense that all the cliques in the junction tree are involved in the process.

3. Belief Updating in Large BNs—New Challenges

There are two concerns regarding belief updating using the GP method for large BNs. First, it is known that belief updating in BNs in general is NP-hard (Cooper 1990). Although it is feasible and even efficient in many real life applications provided the BNs are small. It is foreseeable that larger BNs will take much longer time to perform GP for belief updating. Secondly, the GP method involves inward and outward message passing performed on the *whole* junction tree although not every clique is absolutely necessary to be involved for belief updating (as will be explained shortly.)

4. An On-Demand Thrifty Propagation Method

In this section, we present an on-demand thrifty propagation method based on the GP method. We assume that the GP method is applied in full scale *only once* on a junction tree with no evidence observed and the marginals for cliques and separators in the junction trees are known thereafter.

Consider two adjacent cliques C_i and C_j in a junction tree with the separator $S_{ij} = C_i \cap C_j$. Let $p(C_i)$, $p(C_j)$,

and $p(S_{ij})$ be the marginals corresponding to C_i , C_j , and S_{ij} , respectively. Suppose an evidence \mathbf{E} is observed such that $v(\mathbf{E}) \subseteq C_i$. We then have the following theorem.

Theorem 1 $p(C_j, \mathbf{E}) = p(C_j) \cdot p(\mathbf{E}|S_{ij}) = p(C_j) \cdot \frac{\sum_{C_i - v(\mathbf{E}) - S_{ij}} p(C_i, \mathbf{E})}{p(S_{ij})}$.

Theorem 1 indicates that in order to obtain the updated marginal $p(C_j, \mathbf{E})$, we need to compute $p(\mathbf{E}|S_{ij})$. However, $p(\mathbf{E}|S_{ij}) = \frac{p(\mathbf{E}, S_{ij})}{p(S_{ij})}$, in which the denominator $p(S_{ij})$ is the marginal on S_{ij} and the numerator $p(\mathbf{E}, S_{ij})$ is readily available from $p(C_i)$ (note that $v(\mathbf{E}) \subseteq C_i$) because both $p(C_i)$ and $p(S_{ij})$ has been calculated by the GP method with no evidence observed. We call the computation in the above theorem as that the clique C_i (with the updated marginal $p(C_i, \mathbf{E})$) is passing the evidence \mathbf{E} to clique C_j (to obtain the updated marginal $p(C_j, \mathbf{E})$). It is not hard to see that Theorem 1 implies that clique C_i with its updated probability given evidence \mathbf{E} , i.e., $p(C_i, \mathbf{E})$, can always pass the evidence \mathbf{E} to its adjacent clique C_j to obtained the updated probability $p(C_j, \mathbf{E})$.

Consider a more general case with two adjacent cliques C_i , C_j , and the separator S_{ij} . Let $p(C_i, \mathbf{E}_i)$ and $p(C_j, \mathbf{E}_j)$ be the marginals on C_i and C_j respectively where \mathbf{E}_i and \mathbf{E}_j are compatible evidences. Suppose another evidence \mathbf{E}' compatible with both \mathbf{E}_i and \mathbf{E}_j is observed and $v(\mathbf{E}') \subseteq C_i$. We then have the following theorem.

Theorem 2 The clique C_i with its updated marginal $p(C_i, \mathbf{E}_i, \mathbf{E}')$ passing the evidence \mathbf{E}' to clique C_j results in the updated marginal $p(C_j, \mathbf{E}_j, \mathbf{E}')$.

In many real applications, the queries imposed on a BN exhibit certain patterns as we explain below.

Scenario 1 (multiple variables): Given evidence \mathbf{E} , compute the posterior probability for x_1, x_2, \dots, x_n given \mathbf{E} , namely, $p(x_1|\mathbf{E}), \dots, p(x_n|\mathbf{E})$.

Scenario 2 (multiple evidence): Compute the posterior probability for a variable x given incremental compatible evidences $\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_n$, namely, compute $p(x_1|\mathbf{E}_1), p(x_1|\mathbf{E}_1, \mathbf{E}_2), \dots, p(x_1|\mathbf{E}_1, \dots, \mathbf{E}_n)$.

In the following, we propose a prototype of an on-demand thrifty method for computing posterior probability in the above two scenarios (patterns). We begin our discussion with the simplest case of computing $p(x|\mathbf{E})$ where $v(\mathbf{E})$ is contained by a clique, say C , based on Theorem 1.

PROCEDURE *Compute*(x, \mathbf{E})

Input: variable x and evidence \mathbf{E} such that $v(\mathbf{E})$ is contained by a clique C .

Output: $p(x|\mathbf{E})$.

- 1: Identify a clique C' such that $x \in C'$.
- 2: Find out a path (C_0, C_1, \dots, C_m) in the junction tree such that $C_0 = C, C_m = C'$, and (C_k, C_{k+1}) is an edge in the junction tree, $k = 0, \dots, m-1$.
- 3: For $k = 0$ to $m-1$,
clique C_k passes the evidence \mathbf{E} to C_{k+1} .
- 4: Compute $p(x|\mathbf{E})$ from $p(C', \mathbf{E})$.

5: Mark each clique $C_i, i = 0, \dots, m$, with evidence \mathbf{E} , denoted $C_i^{\mathbf{E}}$.

6: Return $p(x|\mathbf{E})$.

}

Note that the message passing occurring in the above procedure involves only the cliques along the path between C and C' , regardless of whether the junction tree has other cliques. Only the cliques in the path are necessary for the procedure to compute the resulting updated marginal $p(C', \mathbf{E})$, from which $p(x|\mathbf{E})$ can be obtained in Step 4. All the other cliques in the junction tree are irrelevant to the task of the query $p(x|\mathbf{E})$ and they are not involved in computing $p(x|\mathbf{E})$.

A solution to the belief updating scenario 1, namely, the multiple variables case, can be formulated based on the above procedure. Consider a fixed evidence \mathbf{E} and the task of computing $p(x_1|\mathbf{E}), \dots, p(x_n|\mathbf{E})$ for a sequence of variables x_1, \dots, x_n . This can obviously be accomplished by calling the procedure *Compute*(x, \mathbf{E}) with the fixed \mathbf{E} together with different x_1, \dots, x_n variables in the sequence as arguments. However, there is room for improving the efficiency of applying the procedure to a sequence of variables. The scenario of multiple evidence can be solved in a similar fashion based on Theorems 1, 2, and the procedure *Compute*. We can (1) first pick the clique which contains the variable of interest as the root of the junction tree, and (2) whenever new (compatible) evidence \mathbf{E}_i is observed, we pass \mathbf{E}_i from the clique containing the evidence \mathbf{E}_i to the chosen root which contains the variable of interest. The updated marginal on the root thus is always conjoint with all the evidence observed so far.

5. Remarks and Conclusion

The salient feature of the method is that (1) the computation occurred during the propagation is based on the query imposed by the users and it answers the query only, and (2) the computation for obtaining the posterior probability for the variables of interest is minimal in the sense that the propagation only involves those cliques that *have to* participate to produce the results. Such a thrifty method avoids a full scale GP, takes less time to answer queries, and waste less computational resources.

References

- Cooper, G. 1990. The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence* 42(2-3):393-405.
- Jensen, F.; Lauritzen, S.; and Olesen, K. 1990. Bayesian updating in causal probabilistic networks by local computation. *Computational Statistics Quarterly* 4:269-282.
- Lauritzen, S., and Spiegelhalter, D. 1988. Local computation with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society* 50:157-244.
- Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, California: Morgan Kaufmann Publishers.