Proceedings of the Twenty-Second International FLAIRS Conference (2009)

A Comparative Study of Variable Elimination and Arc Reversal in Bayesian Network Inference

C. J. Butz and **J. Chen** and **K. Konkel**
Department of Computer Science
University of Regina
Regina, SK, S4S 0A2
{butz,chen209j,konkel1k}@cs.uregina.ca

P. Lingras
Department of Mathematics and Computing Science
Saint Mary's University
Halifax, NS, B3H 3C3
pawan.lingras@stmarys.ca

Abstract

We compare two approaches to Bayesian network inference, called *variable elimination* (VE) and *arc reversal* (AR). It is established that VE never requires more space than AR, and never requires more computation (multiplications and additions) than AR.

Introduction

Two approaches for eliminating variables from *Bayesian networks* (BNs) (Pearl 1988) are considered here. The first approach, called *variable elimination* (VE) (Zhang and Poole 1994), eliminates a variable by multiplying together all of the distributions involving the variable and then summing the variable out of the obtained product. The second method, known as *arc reversal* (AR) (Olmsted 1983; Shachter 1986), removes a variable v_i with k children in a BN by building $3k - 1$ distributions and outputting k of them. More specifically, three distributions are built when considering each child, except for the last child, when only two distributions need be constructed.

In this paper, we obtain space and computational relationships between VE and AR. Proofs and experimental results will be provided in a separate paper. Note that only rows with positive probability values are stored. We introduce the notion of *row-equivalent* to indicate that the rows (configurations) in one distribution are precisely those in another distribution. Within AR itself, we establish that the rows appearing in the first constructed distribution are exactly those rows appearing in the third constructed distribution, for each child (of the variable being eliminated) except the last. That is, the first distribution built is row-equivalent to the third distribution built for each child except the last. Next, with respect to VE and AR, we show that the only distribution output by VE is row-equivalent to the distribution created for the last child in AR. As AR also requires space for the distributions created for any other children of the variable being eliminated, it is established that VE never requires more space than AR. With respect to computation, we consider multiplication and addition. It is shown that the number of multiplications performed by VE is the same as the number of multiplications required by AR to build the first distribution for every child of the variable being eliminated. However, AR necessarily requires further multiplications when building the third distribution, for all children except the last. Therefore, VE never does more multiplications than AR. Lastly, with respect to addition, the number of additions required by VE for its only marginalization operation are exactly those required by AR to build the second distribution when considering the last child. As AR needs to perform marginalization to build the second distribution for all other children, it is shown that VE never performs more additions than AR.

AR eliminates a variable v_i by reversing the arcs (v_i, v_j) for each child v_j of v_i, where $j = 1, 2, \ldots, k$. With respect to multiplication and addition, AR reverses one arc (v_i, v_j) as a three-step process:

$$p(v_i, v_j | A_j) = p(v_i | A_{j-1}) \cdot p(v_j | P_j), \tag{1}$$

$$p(v_j | B_j) = \sum_{v_i} p(v_i, v_j | A_j), \tag{2}$$

$$p(v_i | A_j) = \frac{p(v_i, v_j | A_j)}{p(v_j | B_j)}. \tag{3}$$

We refer to Eqs. (1) - (3) as the first equation in AR, the second equation in AR, and the third equation in AR, respectively.

For example, to eliminate variable b from the BN shown in Fig. 1, where two of the CPTs are illustrated in Fig. 2, AR reverses arc (b, c) by building the following distributions:

$$p(b, c | a, e) = p(b | a) \cdot p(c | b, e), \tag{4}$$

$$p(c | a, e) = \sum_{b} p(b, c | a, e), \tag{5}$$

$$p(b | a, e, c) = \frac{p(b, c | a, e)}{p(c | a, e)}, \tag{6}$$

as illustrated in Figure 3.

Row Equivalence within AR

When computing the product of two probability distributions, Wong et al. (1995) showed that the rows appearing in the product are precisely the *natural join* (Maier 1983), denoted $\bowtie$, of the two distributions. Similarly, when computing a marginalization of a probability distribution, the rows

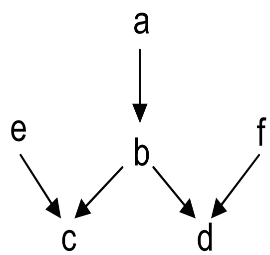

Figure 1: A BN.

a	b	$p(b\|a)$
0	0	0.502
0	1	0.498
1	1	1.000

b	e	c	$p(c\|b,e)$
0	0	0	0.708
0	0	1	0.292
0	1	0	1.000
1	0	0	0.323
1	0	1	0.677
1	1	0	0.358
1	1	1	0.642

Figure 2: CPTs $p(b|a)$ and $p(c|b,e)$ for the BN in Fig. 1.

(i)

a	b	e	c	p(b,c|a,e)
0	0	0	0	0.355
0	0	0	1	0.147
0	0	1	0	0.502
0	1	0	0	0.161
0	1	0	1	0.337
0	1	1	0	0.178
0	1	1	1	0.320
1	1	0	0	0.323
1	1	0	1	0.677
1	1	1	0	0.358
1	1	1	1	0.642

$$\frac{p(b,c\,|\,a,e)}{p(c\,|\,a,e)}$$

(iii)

a	b	e	c	p(b|a,e,c)
0	0	0	0	0.688
0	0	0	1	0.303
0	0	1	0	0.738
0	1	0	0	0.312
0	1	0	1	0.697
0	1	1	0	0.262
0	1	1	1	1.000
1	1	0	0	1.000
1	1	0	1	1.000
1	1	1	0	1.000
1	1	1	1	1.000

$\sum_b$ $\div$

a	e	c	p(c|a,e)
0	0	0	0.516
0	0	1	0.484
0	1	0	0.680
0	1	1	0.320
1	0	0	0.323
1	0	1	0.677
1	1	0	0.358
1	1	1	0.642

(ii)

Figure 3: Row-equivalence of AR Eqs. (4) and (6), namely, the same rows appear in the first and third distributions.

appearing in the marginalization are exactly those defined by the *projection* (Maier 1983), denoted π, of the distribution. This leads us to the following known result in the relational database community.

Theorem 1. *(Maier 1983) Let r be a relation on attributes X and $Y \subseteq X$. Then $r = r \bowtie \pi_Y(r)$.*

In order to obtain the corresponding result for AR, we introduce the notation of *row-equivalent*.

Definition 1. *Given two potentials $\phi_1(X)$ and $\phi_2(X)$ on the same set X of variables, we say $\phi_1(X)$ and $\phi_2(X)$ are row-equivalent, denoted $\phi_1(X) \simeq \phi_2(X)$, if the configurations of X in ϕ_1 are precisely those in ϕ_2.*

In other words, $\phi_1(X) \simeq \phi_2(X)$ means that the same rows appear in both ϕ_1 and ϕ_2. Note that the probability values can still be different.

Theorem 2. *Let $\phi(X)$ be a potential and $Y \subseteq X$. Let*

$$\phi'(X) \quad = \quad \frac{\phi(X)}{\sum_Y \phi(X)}.$$

Then $\phi(X) \simeq \phi'(X)$.

The rows in $p(v_i, v_j|A_j)$ of the first equation in AR can now be shown to be precisely those in $p(v_i|A_j)$ of the third equation in AR.

Theorem 3. *Suppose AR will eliminate variable v_i with k children $v_1, v_2, \ldots, v_k$ in a given BN. Then the distribution created in Eq. (1) is row-equivalent to the distribution created in Eq. (3), for $j = 1, 2, \ldots, k-1$.*

For example, recall the three distributions built in Eqs. (4)-(6) and that are depicted in Fig. 3. It can be seen that the rows in distribution $p(b, c|a, e)$ of Eq. (4) are exactly the same rows in distribution $p(b|a, e, c)$ of Eq. (6).

Relationships Between VE and AR

The establishment of row equivalence within AR allows us to comment on the space relationship between AR and VE.

Theorem 4. *VE never requires more space than AR to eliminate a set of variables from a BN.*

Since VE never uses more space than AR, it can be shown that VE does less work than AR.

Theorem 5. *VE never requires more multiplications, nor more additions, than AR to eliminate a set of variables from a BN.*

References

Maier, D. 1983. *The Theory of Relational Database*. Computer Science Press.

Olmsted, S. 1983. *On representing and solving decision problems*. Ph.D. Dissertation, Stanford University, Department of Engineering Economic Systems.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Shachter, R. 1986. Evaluating influence diagrams. *Operations Research* 34(6):871–882.

Wong, S.; Butz, C.; and Xiang, Y. 1995. A method for implementing a probabilistic model as a relational database. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 556–564.

Zhang, N., and Poole, D. 1994. A Simple Approach to Bayesian Network Computations. In *Proceedings of the Tenth Canadian Conference on Artificial Intelligence*, 171–178.