Incorporating Evidence in Bayesian networks with the Select Operator

C.J. Butz and F. Fang

Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada SAS 0A2
{butz, fang11fa}@cs.uregina.ca

Abstract. In this paper, we propose that the select operator in relational databases be adopted for incorporating evidence in Bayesian networks. This approach does not involve the construction of new evidence potentials, nor the associated computational costs of multiplying the evidence potentials into the knowledge base. The select operator also provides unified treatment of hard and soft evidence in Bayesian networks. Finally, some query optimization rules, involving the select operator implemented in relational databases, can be directly incorporated into probabilistic expert systems.

1 Introduction

Bayesian networks [3] are an established framework for uncertainty management and have been successfully applied in practice in a variety of problem domains. Processing evidence is a fundamental task in Bayesian networks [2]. Evidence means that some information about the values of a set E of variables is obtained. For instance, the exact values $E = e$ are known or perhaps that E does not take specific values $E \neq e$. Given evidence, say $E = e$, several techniques have been proposed for processing queries of the form $p(X \mid E = e)$, where X is a set of non-evidence variables. These techniques, however, have two disadvantages. First, the *construction* of a new probability table, called an *evidence potential*, is required. Second, the evidence potential is *multiplied* with the probability tables stored in the knowledge.

Several researchers, including [4, 5], have pointed out the intrinsic relationship between Bayesian networks and relational databases [1]. More recently, Wong et al. [5] established that the logical implication of probabilistic conditional independence exactly coincides with that of embedded multivalued dependency in relational databases for the classes of Bayesian networks, Markov networks, and fixed-context.

In this paper, we propose that the *select operator* [1] in relational databases be adopted for incorporating evidence in Bayesian networks. We first show that the select operator can be applied on probabilistic relations and not only traditional relations. As the name suggests, the select operator selects a subset of rows from a probabilistic relation that satisfy the select condition. Our main result is that incorporating evidence with the select operator is equivalent to the approach with evidence potentials. Our approach does *not* involve the construction of new evidence potentials, nor

the associated computational costs of *multiplying* the evidence potentials into the knowledge base. The select operator also provides *unified* treatment of hard and soft evidence in Bayesian networks. Finally, some query *optimization rules*, involving the select operator implemented in relational databases, can be directly incorporated into probabilistic expert systems.

This paper is organized as follows. The select operator and query optimization rules are presented in Section 2. In Section 3, we discuss two methods for incorporating evidence in Bayesian networks. In Section 4, advantages of processing evidence with the select operator are provided. The conclusion is presented in Section 5.

2 The Select Operator

Wong et al. [4, 5] have shown how probability tables in Bayesian networks can be viewed as *probabilistic relations*, i.e., traditional relations in conventional databases [1] can be extended with a probability column. In this section, we incorporate the *select operator* [1] into this probabilistic setting.

Let $r(X)$ be a probabilistic relation on X. Let A be an attribute in X, $a \in \text{dom}(A)$, and *op* be an operator in $\{=, \neq, <, \leq, \geq, >\}$. Then

$$\sigma_{A \; op \; a}(r) = \{ \; t \mid t \in r \text{ and } t(A) \; op \; a \; \}, \tag{1}$$

where $t(A)$ is the *restriction* of tuple t to attribute A.

Example 1. In Table 1, given the probabilistic relation $r(z|h)$ on the left, $\sigma_{z=1}(\; r(z|h)\;)$ and $\sigma_{z \neq 0}(\; r(z|h)\;)$ are shown in the middle and on the right, respectively.

Table 1. A probabilistic relation $r(z|h)$ (left). $\sigma_{z=1}(\; r(z|h)\;)$ (middle). $\sigma_{z \neq 0}(\; r(z|h)\;)$ (right)

h	z	$p(z|h)$		h	z	$p(z=1|h)$		h	z	$p(z \neq 0|h)$
1	2	0.123		1	1	0.456		1	2	0.123
1	1	0.456		0	1	0.333		1	1	0.456
1	0	0.421						0	2	0.111
0	2	0.111						0	1	0.333
0	1	0.333								
0	0	0.556								

We now consider the relationship of the select operator with the multiplication and marginalization operators implemented in probabilistic expert systems. Here X and Y are sets of attributes and the select conditions do not involve probability columns. The soundness of these rules follows from the corresponding rules in databases [1].

1: The σ operator is commutative. Given select conditions c_1 and c_2,

$$\sigma_{c_1}(\; \sigma_{c_2}(\; r(X)\;)\;) \quad = \quad \sigma_{c_2}(\; \sigma_{c_1}(\; r(X)\;)\;). \tag{2}$$

2: Given a conjunction of select conditions $\{\; c_1, c_2, ..., c_m \;\}$ involving attributes in X:

$$\sigma_{c_1 \wedge c_2 \wedge ... \wedge c_m}(\; r(X)\;) \quad = \quad \sigma_{c_1}(\; \sigma_{c_2} ...(\; \sigma_{c_m}(\; r(X)\;)\;)\;). \tag{3}$$

3: If the select condition c only involves attributes in X, then

$$\Sigma_X \left(\sigma_c \left(r(XY) \right) \right) \quad = \quad \sigma_c \left(\Sigma_X \left(r(XY) \right) \right). \tag{4}$$

4: If the select condition c only involves attributes in X, then

$$\sigma_c \left(r(Y) \otimes r(X) \right) \quad = \quad r(Y) \otimes \sigma_c \left(r(X) \right), \tag{5}$$

where $\otimes$ is a multiplication join for probabilistic relations [4].

3 Incorporating Evidence in Bayesian networks

Traditionally, evidence in incorporated into Bayesian networks [3] using *evidence potentials* [2]. We suggest using the select operator.

3.1 With Evidence Potentials

In the literature [2], there are two distinct kinds of evidence, namely, hard and soft.

Hard evidence is an instantiation of a set E of variables, i.e., it is observed that $E = e$. Hard evidence is incorporated into a Bayesian network as follows. First, a *finding potential* $F(E = e)$ is constructed. The probability column of $F(E = e)$ is set as follows: for the row with $E = e$, the probability value is one; for all other rows, the probability value is zero. Second, $F(E = e)$ is multiplied with the stored CPTs.

Soft evidence means $E \neq e$, i.e., it is known that the set E of variables does not take on value e. Soft evidence is incorporated into a Bayesian network as follows. First, a *likelihood potential* $L(E \neq e)$ is constructed. The probability column of $L(E \neq e)$ is set as follows: for the rows with $E \neq e$, the probability value is one; for all other rows, the probability value is zero. Second, $L(E \neq e)$ is multiplied with the stored CPTs.

Example 2. Suppose we observe hard evidence $z = 1$. The constructed finding potential $F(z = 1)$ is shown in Table 2 (left). The product $F(z = 1) \cdot p(z|h)$ is illustrated in Table 2, where $p(z|h)$ is the CPT from Table 1 (left). Similarly, given soft evidence $z \neq 0$, the CPT $p(z|h)$ is multiplied with the constructed likelihood potential $L(z \neq 0)$.

Table 2. Given hard evidence $z = 1$, the finding potential $F(z=1)$ is multiplied with the CPT $p(z|h)$ giving $p(z=1|h)$

z	F(z=1)		h	z	p(z|h)		h	z	p(z=1|h)
0	0	.	1	2	0.123	=	1	1	0.456
1	1		1	1	0.456		0	1	0.333
2	0		1	0	0.421				
			0	2	0.111				
			0	1	0.333				
			0	0	0.556				

3.2 With the Select Operator

Here, we incorporate evidence in Bayesian networks using the select operator.

Consider hard evidence $A_i = a_i$, where $a_i \in \mathrm{dom}(A_i)$. The select condition c is the hard evidence $A_i = a_i$. Compute $\sigma_{A_i = a_i}(\, r(A_i|P_i)\,)$, where $r(A_i|P_i)$ is the probabilistic relation for the CPT $p(A_i|P_i)$ of the hard evidence variable A_i.

Consider soft evidence $A_i \neq a_i$, where $a_i \in \mathrm{dom}(A_i)$. The select condition c is the soft evidence $A_i \neq a_i$. Compute $\sigma_{A_i \neq a_i}(\, r(A_i|P_i)\,)$, where $r(A_i|P_i)$ is the probabilistic relation for the CPT $p(A_i|P_i)$ of the hard evidence variable A_i.

Example 3. Recall the probabilistic relation $r(z|h)$ in Table 1 (left). The hard evidence $z = 1$ is incorporated using the select operator as $\sigma_{z=1}(\, r(z|h)\,)$, as shown in Table 1 (middle). Similarly, the soft evidence $z \neq 0$ is incorporated using the select operator as $\sigma_{z\neq0}(\, r(z|h)\,)$, as shown in Table 1 (right).

Theorem 1. Given a Bayesian network, suppose hard evidence $A_i = a_i$ is observed. Incorporating $A_i = a_i$ with the select operator is equivalent to using a finding potential.

Proof: Consider the constructed finding potential $F(A_i = a_i)$. There is one row in table $F(A_i = a_i)$ for each value in $\mathrm{dom}(A_i)$. The probability column in $F(A_i = a_i)$ is set as follows: one, for the row with $A_i = a_i$; zero, otherwise. Now consider the product of the finding potential $F(A_i = a_i)$ with the Bayesian network CPT $p(A_i|P_i)$ for variable A_i. By definition, rows of $F(A_i = a_i)$ will be multiplied with rows $p(A_i|P_i)$ provided they have the same value for A_i. Since the probability value in $F(A_i = a_i)$ is zero for all rows with $A_i \neq a_i$, the only rows appearing in the product $F(A_i = a_i) \cdot p(A_i|P_i)$ are those with $A_i = a_i$. Moreover, since the probability column of $F(A_i = a_i)$ is one when $A_i = a_i$, the probabilities of $p(A_i|P_i)$ with $A_i = a_i$ and $F(A_i = a_i) \cdot p(A_i|P_i)$ are equal. Hence, the result is the selection of those rows of $p(A_i|P_i)$ with $A_i = a_i$. This is the definition of $\sigma_{A_i = a_i}(\, r(A_i|P_i)\,) = \{\, t \mid t \in r(A_i|P_i) \text{ and } t(A_i) = a_i \,\}$, where $r(A_i|P_i)$ is the probabilistic relation for the CPT $p(A_i|P_i)$ of the hard evidence variable A_i.

Corollary 1. Given a Bayesian network, suppose soft evidence $A_i \neq a_i$ is observed. Incorporating $A_i \neq a_i$ with the select operator is the same as using a likelihood potential.

For instance, the hard evidence $z = 1$ can be incorporated either using the select operator as in Table 1 (middle) or with a finding potential as in Table 2.

4   Advantages of Processing Evidence with the Select Operator

In this section, we give three advantages of using the select operator to incorporate evidence in Bayesian networks, namely, (i) there is no need to *construct* an evidence

potential, (ii) we can *reduce* the number of multiplications, and (iii) the select operator *unifies* the notions of hard evidence and soft evidence.

The select operator does *not* require the construction of evidence potentials. On the contrary, the traditional approach [2] requires the construction of finding potentials for hard evidence, and likelihood potentials for soft evidence. See, for instance, $F(z = 1)$ in Table 2. It is well known that query optimization in Bayesian networks involves *reducing* the number of multiplications. While the traditional approach necessarily involves some multiplications to incorporate the evidence potentials into the knowledge base (i.e., Table 2), the select operator does not involve constructing evidence tables, nor their associated multiplicative costs (i.e., Table 1 (middle)). The last advantage concerns the *distinct* treatment of hard and soft evidence in the traditional approach. Here hard evidence is incorporated with *finding potentials*, while soft evidence involves *likelihood potentials*. On the contrary, the select operator readily handles both hard and soft evidence.

5   Conclusion

One objective in query optimization is to reduce the number of multiplications [2]. On the contrary, the traditional approach to incorporate evidence in Bayesian networks constructs new evidence tables simply to multiply them into the knowledge base (see Table 2). By adopting the select operator from relational databases, evidence can be incorporated without constructing evidence tables and without their associated multiplicative costs (see Table 1 (middle)). Moreover, the select operator provides unified treatment of both hard and soft evidence. We have also shown how select interacts with the multiplication and marginalization operators in probabilistic expert systems. Thus, the above analysis suggests that the select operator is a very convenient tool for incorporating evidence in Bayesian networks.

Reference

1. Abiteboul, S., Hull, R., and Vianu, V.: Foundations of Databases. Addison-Wesley Publishers, United States of America (1995)
2. Madsen, A.L. and Jensen, F.V.: Lazy Propagation: A Junction Tree Inference Algorithm Based on Lazy Evaluation. Artificial Intelligence 113 (1-2) (1999) 203-245
3. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, San Francisco, California (1988)
4. Wong, S.K.M., Butz, C.J. and Xiang, Y.: A Method for Implementing a Probabilistic Model as a Relational Database. 11th Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers, Montreal, Quebec (1995) 556-564
5. Wong, S.K.M., Butz, C.J. and Wu, D.: On the Implication Problem for Probabilistic Conditional Independency. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, Vol. 30, No. 6, (2000) 785-805