A Method for Detecting Context-Specific Independence in Conditional Probability Tables

C.J. Butz and M.J. Sanscartier

Department of Computer Science, University of Regina,
Regina, Canada, S4S 0A2, {butz,sanscarm}@cs.uregina.ca

Abstract. Context-specific independence is useful as it can lead to improved inference in Bayesian networks. In this paper, we present a method for detecting this kind of independence from data and emphasize why such an algorithm is needed.

1 Introduction

Based upon the notions of *conditional probability tables* (CPTs) and *probabilistic conditional independence* [3], the *Bayesian network* [2] is an elegant and formal framework for probabilistic reasoning. More recently, the Bayesian community has become interested in *contextual* independencies such as *context-specific independence* (CSI) [1]. Contextual independencies are useful since they may lead to more efficient probabilistic inference [4]. However, the *acquisition* of contextual independencies has not received as much attention. In [1], *CPT-trees* were introduced to help a human expert specify a CPT. A graphical method, which we call *csi-detection*, was provided to read CSIs from a CPT-tree [1].

In some situations, however, no human expert is available. In addition, we explicitly demonstrate that the csi-detection may *fail* to detect valid CSIs holding in the CPT-tree constructed directly from a given CPT. Thus, a method for detecting CSIs from data is needed. In this paper, we suggest a procedure (Algorithm 1) for detecting CSIs in a given CPT.

This paper is organized as follows. Section 2 introduces *context-specific independence*. In Section 3, we review a method for obtaining CSIs from an expert. In Section 4, we propose a method for detecting CSIs in a given CPT. The conclusion is presented in Section 5.

2 Context-Specific Independence

Let p be a *joint probability distribution* (jpd) [3] over a set U of variables and X, Y, Z be subsets of U. We say Y and Z are *conditionally independent* given X, if given any $x \in V_X$, $y \in V_Y$, then for all $z \in V_Z$,

$$p(y \mid x, z) \quad = \quad p(y \mid x), \qquad \text{whenever } p(x, z) > 0. \tag{1}$$

Consider a Bayesian network with directed edges $\{(A,C), (A,D), (A,E), (B,D), (C,E), (D,E)\}$. Based on the *conditional independence* (CI) assumptions encoded in this network, the jpd $p(A, B, C, D, E)$ can be factorized as

$$p(A, B, C, D, E) \quad = \quad p(A) \cdot p(B) \cdot p(C|A) \cdot p(D|A, B) \cdot p(E|A, C, D), \qquad (2)$$

where $p(D|A, B)$ and $p(E|A, C, D)$ are shown in Fig. 1. The marginal $p(A, B, C, E)$ can be computed from Eq. (2) as follows: (i) compute the product $p(D|A, B) \cdot p(E|A, C, D)$; (ii) marginalize out variable D from this product; and (iii) multiply the resulting distribution with $p(A) \cdot p(B) \cdot p(C|A)$.

A	B	D	$p(D\|A, B)$
0	0	0	0.3
0	0	1	0.7
0	1	0	0.3
0	1	1	0.7
1	0	0	0.6
1	0	1	0.4
1	1	0	0.8
1	1	1	0.2

A	C	D	E	$p(E\|A, C, D)$
0	0	0	0	0.1
0	0	0	1	0.9
0	0	1	0	0.1
0	0	1	1	0.9
0	1	0	0	0.8
0	1	0	1	0.2
0	1	1	0	0.8
0	1	1	1	0.2
1	0	0	0	0.6
1	0	0	1	0.4
1	0	1	0	0.3
1	0	1	1	0.7
1	1	0	0	0.6
1	1	0	1	0.4
1	1	1	0	0.3
1	1	1	1	0.7

Fig. 1. The CPTs $p(D|A, B)$ and $p(E|A, C, D)$ in Eq. (2).

In some situations, however, the conditional independence may only hold for certain *specific* values in V_X, called *context-specific independence* (CSI) [1]. Let X, Y, Z, C be pairwise disjoint subsets of U and $c \in V_C$. We say Y and Z are *conditionally independent* given X in *context* $C = c$, if

$$p(y \mid x, z, c) = p(y \mid x, c), \qquad \text{whenever } p(x, z, c) > 0.$$

For example, consider again the CPT $p(D|A, B)$ redrawn in Fig. 2 (i). Although variables D and B are *not* conditionally independent given A, it can be seen in Fig. 2 (ii,iii) that D and B are independent in context $A = 0$. Similarly, for the CPT $p(E|A, C, D)$, variables E and D are independent given C in context $A = 0$, while variables E and C are independent given D in context $A = 1$.

The CPTs $p(D|A, B)$ and $p(E|A, C, D)$ can then be rewritten as

$$p(D|A, B) \quad = \quad p(D|A = 0) \odot p(D|A = 1, B), \qquad (3)$$

and

$$p(E|A, C, D) \quad = \quad p(E|A = 0, C) \odot p(E|A = 1, D), \qquad (4)$$

A	B	D	$p(D\|A,B)$
0	0	0	0.3
0	0	1	0.7
0	1	0	0.3
0	1	1	0.7
1	0	0	0.6
1	0	1	0.4
1	1	0	0.8
1	1	1	0.2

(i)

A	B	D	$p(D\|A=0,B)$
0	0	0	0.3
0	0	1	0.7
0	1	0	0.3
0	1	1	0.7

A	B	D	$p(D\|A=1,B)$
1	0	0	0.6
1	0	1	0.4
1	1	0	0.8
1	1	1	0.2

(ii)

A	D	$p(D\|A=0)$
0	0	0.3
0	1	0.7

(iii)

Fig. 2. Variables D and B are conditionally independent in context $A = 0$.

where $\odot$ is the *union product* operator [4]. By substituting Eqs. (3) and (4) into Eq. (2), the factorization of the jpd $p(A, B, C, D, E)$ using CSI is

$$p(A,B,C,D,E) = p(A) \cdot p(B) \cdot p(C|A) \odot p(D|A=0) \odot p(D|A=1,B) \odot p(E|A=0,C) \odot p(E|A=1,D). \quad (5)$$

Computing $p(A, B, C, E)$ from Eq. (5) requires 16 fewer multiplications and 8 fewer additions compared to the respective number of computations needed to compute $p(A, B, C, E)$ from the CI factorization in Eq. (2).

3 Specification of CSIs by a Human Expert

Instead of viewing a CPT as a table, here we view a CPT as a tree structure, called a *CPT-tree* [1]. The CPT-tree representation is advantageous since it makes it particularly easy to elicit probabilities from a human expert. A second advantage of CPT-trees is that they allow a simple graphical method, which we call *csi-detection*, for detecting CSIs [1]. We describe csi-detection as follows.

Given a CPT-tree for a variable A and its parent set π_A, i.e., a CPT-tree for the CPT $p(A|\pi_A)$. The *label* of a path is defined as the value of the nodes on that path. A path is *consistent* with a context $C = c$ iff the labeling of the path is consistent with the assignment of the values in c. Given the CPT-tree depicting $p(Y|X, Z, C)$, we say that variable Y is independent of variable Z given X in the specific context $C = c$, if Z does not appear on any path consistent with $C = c$.

Example 1. A human expert could specify the the CPT-tree in Fig. 3 representing the CPT $p(E|A, C, D)$ in Fig. 1. Consider the context $A = 0$. Since variable D does not appear on any path consistent with $A = 0$, we say that variables E and D are independent given C in context $A = 0$. It can be verified that variables E and C are independent given D in context $A = 1$.

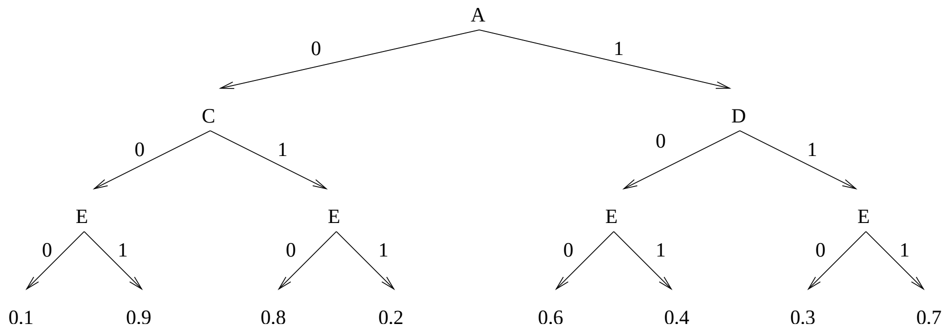

Fig. 3. The CPT-tree given by a human expert representing $p(E|A, C, D)$ in Fig. 1.

4 Detecting CSIs in a Conditional Probability Table

In this section, we propose a method for detecting context-specific independencies from a CPT. We begin by showing why this approach is needed.

In many situations, no human expert is available and one must rely solely on data. Moreover, the csi-detection method presented in the last section may *not* work on the CPT-tree built directly from a given CPT.

Example 2. Suppose there is no human expert available. The *initial* CPT-tree in Fig. 4 is obtained directly from the CPT in Fig. 1. Although variables E and D are independent given C in context $A = 0$, while variables E and C are independent given D in context $A = 1$, the csi-detection method does *not* detect any CSIs holding in this initial CPT-tree.

The problem here is that the csi-detection method is based on missing arcs in the CPT-tree. Thus, we suggest the following algorithm to remove the vacuous arcs in the initial CPT-tree constructed directly from a given CPT.

Algorithm 1 REFINED CPT-TREE
Input: an *initial* CPT-tree for a given CPT
Output: the *refined* CPT-tree obtained by removing all vacuous arcs
1. If all children of a node A are identical, then replace A by one of its offspring.
2. Delete all other children of node A.

Example 3. Consider again the initial CPT-tree in Fig. 4. When $A = 0$ and $C = 0$, node D has identical children. Hence, node D can be replaced with node E. Similarly, for when $A = 0$ and $C = 1$. Moreover, when $A = 1$, node C has identical children. Node C can then be replaced by node D. The *refined* CPT-tree after these deletions is shown in Fig. 3.

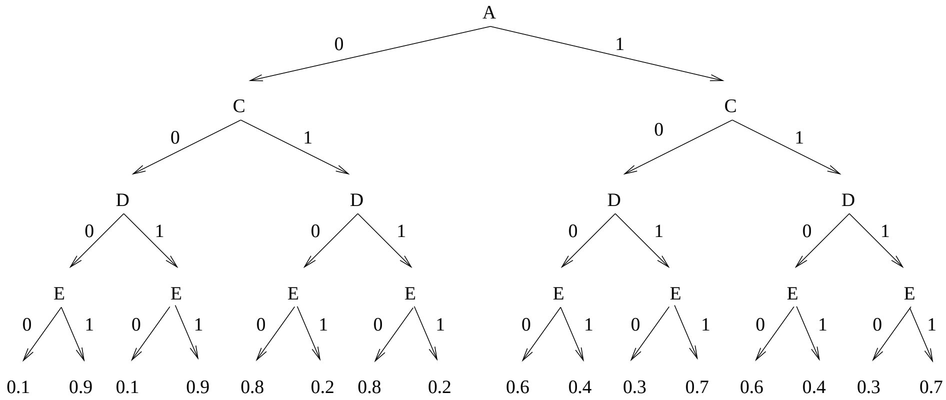

Fig. 4. The *initial* CPT-tree for the given CPT $p(E|A,C,D)$ in Fig. 1.

5 Conclusion

Contextual independencies such as *context-specific independence* (CSI) [1] are important, since they can lead to more efficient inference [4]. Previous work has suggested using *CPT-trees* and *csi-detection* to elicit CSIs from an expert [1]. In some situations, however, no human expert is available. Moreover, Example 2 explicitly demonstrates that the csi-detection may *fail* to detect valid CSIs holding in the CPT-tree constructed directly from a given CPT. Thus, a method for detecting CSIs from data is needed. In this paper, we proposed Algorithm 1 for detecting CSIs in a given CPT.

References

1. Boutilier, C., Friedman, N., Goldszmidt, M., Koller, D.: Context-specific independence in Bayesian networks, *Twelfth Conference on Uncertainty in Artificial Intelligence*, 115–123, 1996.
2. Pearl, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, 1988.
3. Wong, S.K.M., Butz, C.J., Wu, D.: On the implication problem for probabilistic conditional independency. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 30, Part A, No. 6, 785–805, 2000.
4. Zhang, N., Poole, D.: On the role of context-specific independence in probabilistic inference, *Sixteenth International Joint Conference on Artificial Intelligence*, 1288–1293, 1999.

This article was processed using the LaTeX macro package with LLNCS style