Acquisition Methods for Contextual Weak Independence

C.J. Butz and M.J. Sanscartier

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

Abstract. Although *contextual weak independence* (CWI) has shown promise in leading to more efficient probabilistic inference, no investigation has examined how CWIs can be obtained. In this paper, we suggest and analyze two methods for obtaining this kind of independence.

1 Introduction

Probabilistic reasoning would not be feasible without making some kind of independency assumptions. By making *conditional independence* (CI) assumptions, *Bayesian networks* have become an established framework for uncertainty management. More recently, focus has shifted somewhat from *non-contextual* independencies such as CI to *contextual* independencies such as *contextual weak independence* (CWI) [3] and *context-specific independence* (CSI) [1]. Although more efficient probabilistic inference can be achieved in a CWI approach using independencies that would go unnoticed in a CSI approach [2], *no* study has ever investigated how CWIs can be obtained.

In this paper, we suggest two methods for obtaining contextual weak independencies. The first approach is to directly obtain the independencies from a human expert. In the case when no expert is available, the second approach we propose is to detect the CWIs from data. This investigation, by giving one method to obtain CWIs from an expert and another for obtaining CWIs from data, complements the work in [2] on more efficient inference using CWIs.

This paper is organized as follows. Section 2 introduces *contextual weak independence* (CWI). In Section 3, we give a method for obtaining CWIs from an expert. In Section 4, we propose a method for detecting CWIs in a given CPT. The conclusion is presented in Section 5.

2 Contextual Weak Independence

Consider the *conditional probability table* (CPT) $p(D|A, B)$ in Fig. 1 (i). It can be verified that variables D and B are *not* conditionally independent given A. The notion of *context-specific independence* (CSI) allows for the case when the independence holds in a particular context of A. It can be verified that variables D and B are *not* conditionally independent in the context $A = 0$, nor in context $A = 1$. The important point is that CI and CSI *fail* to simplify $p(D|A, B)$.

A	B	D	$p(D\|A,B)$
0	0	0	0.3
0	0	1	0.1
0	0	2	0.6
0	1	0	0.3
0	1	1	0.1
0	1	2	0.6
0	2	0	0.3
0	2	1	0.1
0	2	2	0.6
0	3	3	1.0
1	0	0	0.2
1	0	1	0.8
1	1	0	0.2
1	1	1	0.8
1	2	2	0.4
1	2	3	0.6
1	3	2	0.4
1	3	3	0.6

(i)

A	B	D	$p(D\|A=0,B)$
0	0	0	0.3
0	0	1	0.1
0	0	2	0.6
0	1	0	0.3
0	1	1	0.1
0	1	2	0.6
0	2	0	0.3
0	2	1	0.1
0	2	2	0.6
0	3	3	1.0

A	B	D	$p(D\|A=1,B)$
1	0	0	0.2
1	0	1	0.8
1	1	0	0.2
1	1	1	0.8
1	2	2	0.4
1	2	3	0.6
1	3	2	0.4
1	3	3	0.6

(ii)

A	D	$p_1(D\|A=0)$
0	0	0.3
0	1	0.1
0	2	0.6

A	B	D	$p_2(D\|A=0,B)$
0	3	3	1.0

A	D	$p_3(D\|A=1)$
1	0	0.2
1	1	0.8

A	D	$p_4(D\|A=1)$
1	2	0.4
1	3	0.6

(iii)

Fig. 1. Unlike CI and CSI, the notion of CWI can simplify the CPT $p(D|A,B)$.

Let U be a finite set of variables and V_X be the frame of $X \subseteq U$. Let X, Y, Z, C be pairwise disjoint subsets of U and $c \in V_C$. We say Y and Z are *weakly independent* [3] given X in context $C = c$, if both of the following two conditions are satisfied: (i) there exists a maximal disjoint compatibility class $\pi = \{t_i, \ldots, t_j\}$ in the relation $\theta(X, Y, C = c) \circ \theta(X, Z, C = c)$, and (ii) given any $x \in V_X^{\pi}$, $y \in V_Y^{\pi}$, then for all $z \in V_Z^{\pi}$,

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

where $\theta(W)$ denotes the equivalence relation induced by the set W of variables, $\circ$ denotes the composition operator, and V_W^{π} denotes the set of values for W appearing in π.

Example 1. Let us partition the CPT $p(D|A,B)$ in Fig. 1 (i) into the four blocks shown in Fig. 1 (ii). In three of these blocks, variables D and B are conditionally independent given A. Hence, variable B can be dropped as shown in Fig. 1 (iii). For simplicity, $p_{B\in\{0,1,2\}}(D|A=0)$, $p(D|A=0, B=3)$, $p_{B\in\{0,1\}}(D|A=1)$, and $p_{B\in\{2,3\}}(D|A=1)$, are written as $p_1(D|A=0)$, $p_2(D|A=0,B)$, $p_3(D|A=1)$, and $p_4(D|A=1)$, respectively.

In practice, the four *partial functions* $p_1(D|A=0)$, $p_2(D|A=0,B)$, $p_3(D|A=1)$, and $p_4(D|A=1)$ will be stored in place of the single CPT $p(D|A,B)$. Unfortunately, the *union product* operator $\odot$ in [4] for combining partial functions is not sufficient for CWI, since

$$p(D|A,B) \neq p_1(D|A=0) \odot p_2(D|A=0,B) \odot p_3(D|A=1) \odot p_4(D|A=1).$$

By $\oplus$, we denote the *weak join* operator defined as:

$$p(y,x) \oplus q(x,z) = \begin{cases} p(y,x) \cdot q(x,z) & \text{if both } p(y,x) \text{ and } q(x,z) \text{ are defined} \\ p(y,x) & \text{if } p(y,x) \text{ is defined and } q(x,z) \text{ is undefined} \\ q(x,z) & \text{if } p(y,x) \text{ is undefined and } q(x,z) \text{ is defined} \\ 0.0 & \text{if } p(y,x) \text{ and } q(x,z) \text{ are inconsistent} \\ \text{undefined} & \text{if both } p(y,x) \text{ and } q(x,z) \text{ are undefined.} \end{cases}$$

We leave it to the reader to verify that:

$$p(D|A,B) = p_1(D|A=0) \oplus p_2(D|A=0,B) \oplus p_3(D|A=1) \oplus p_4(D|A=1).$$

The important point in this section is that CWI together with the weak join operator $\oplus$ allow the given CPT $p(D|A,B)$ to be faithfully represented by four *smaller* distributions. Such a CWI representation can lead to more efficient inference than with CI and CSI [2]. Thereby, it is useful to study ways to obtain the CWIs holding in a problem domain.

3 Specification of CWIs 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. The *label* of a path in a CPT-tree is defined as the value of the nodes on that path. By (A,B,X) we denote a directed edge from variable A to variable B in the CPT-tree with label X.

Example 2. A human expert could specify the CPT-tree in Fig. 2 representing the CPT $p(D|A,B)$ in Fig. 1.

Algorithm 1 transforms a single CPT into several *partial* CPTs based on the CWIs in the CPT-tree given by a human expert.

Algorithm 1 DECOMPOSE CPT
Input: an expert specified CPT-tree defining a single CPT $p(A|X)$
Output: the *partial* CPTs faithfully representing $p(A|X)$
1. Block the CPT according to the levels in the CPT-tree.
2. For any edge (A,B,X) where X is non-singleton,
Remove A from the corresponding block.

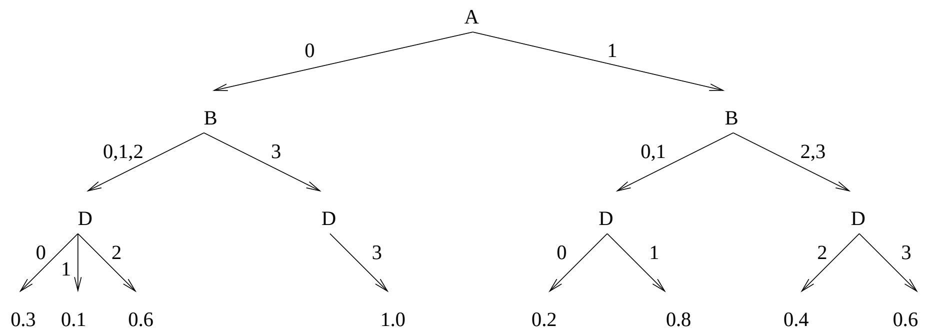

Fig. 2. The CPT-tree given by a human expert representing $p(D|A,B)$ in Fig. 1.

Example 3. The given expert specified CPT-tree in Fig. 2 defines the CPT $p(D|A,B)$ in Fig. 1 (i). By step (1) of Algorithm 1, the CPT-tree indicates the following four blocks:

$$\begin{aligned} &Block\ \#1: \quad A \in \{0\}, B \in \{0,1,2\}, D \in \{0,1,2\},\\ &Block\ \#2: \quad A \in \{0\}, B \in \{3\}, D \in \{3\},\\ &Block\ \#3: \quad A \in \{1\}, B \in \{0,1\}, D \in \{0,1\},\\ &Block\ \#4: \quad A \in \{1\}, B \in \{2,3\}, D \in \{2,3\}. \end{aligned}$$

By step (2) of Algorithm 1, variable B can be deleted from the first, third, and fourth partial CPTs. Observe that this indicates the presence of CWIs, for instance, $p_1(D|A=0,B) = p_1(D|A=0)$.

4 Detecting CWIs in a Conditional Probability Table

In this section, we propose a method for detecting contextual weak independencies from a CPT. Such a method is useful when no human expert is available.

As it may be difficult to determine CWIs directly from a CPT, a given CPT can be represented as a CPT-tree. For example, given the CPT $p(D|A,B)$ in Fig. 1 (i), one *initial* CPT-tree is shown in Fig. 3. Algorithm 2 *refines* an initial CPT-tree so that Algorithm 1 can be applied.

Algorithm 2 REFINE CPT-TREE
Input: an *initial* CPT-tree for a given CPT
Output: the *refined* CPT-tree obtained by removing all vacuous edges
If some children of a node A are identical, then combine these children into one node by augmenting the labels of the combined edges.

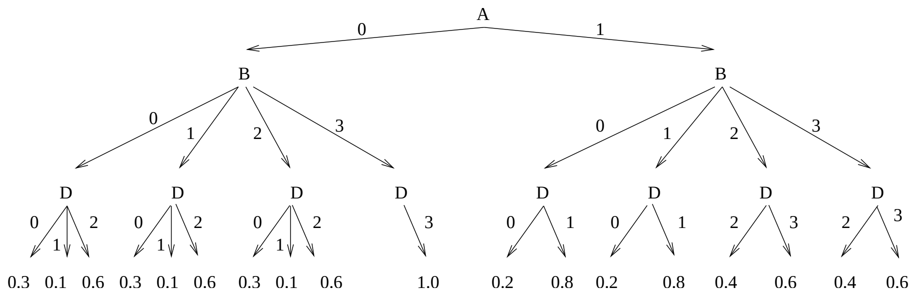

Fig. 3. One initial CPT-tree defined by the CPT $p(D|A, B)$ in Fig. 1.

Example 4. Consider the initial CPT-tree in Fig. 3. When $A = 0$, node B has identical children, namely, the three children with edge labels 0, 1, and 2. Hence, these three edges can be combined into a single edge with label $0, 1, 2$. Moreover, when $A = 1$, node B has identical children for its values 0 and 1. Hence, these edges are grouped together. Similarly, for when $A = 1$ and $B \in \{2, 3\}$. The *refined* CPT-tree after these deletions is shown in Fig. 2.

The important point is that CWIs can be obtained from a CPT. From a given CPT, we can construct an *initial* CPT-tree. Algorithm 2 can then be applied to *refine* the tree. Finally Algorithm 1 can be applied to decompose the given CPT into the smaller *partial* CPTs based on detected CWIs.

5 Conclusion

We have suggested a method (Algorithm 1) for obtaining CWIs from a human expert and another (Algorithm 2) for the situation when no expert is available. Acquiring CWIs is quite important since probabilistic inference using CWIs can be more efficient than inference using CIs and CSIs [2].

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. Butz, C.J., Sanscartier, M.J.: On the role of contextual weak independence in probabilistic inference, *Fifteenth Canadian Conference on Uncertainty in Artificial Intelligence*, (to appear) 2002.
3. Wong, S.K.M., Butz, C.J.: Contextual weak independence in Bayesian networks. In *Fifteenth Conference on Uncertainty in Artificial Intelligence*, 670–679, 1999.
4. Zhang, N., Poole, D.: On the role of context-specific independence in probabilistic inference, *Sixteenth Int. Joint Conf. on Artificial Intelligence*, 1288–1293, 1999.

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