Properties of Weak Conditional 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. *Object-oriented Bayesian networks* (OOBNs) facilitate the design of large Bayesian networks by allowing Bayesian networks to be nested inside of one another. *Weak conditional independence* has been shown to be a necessary and sufficient condition for ensuring consistency in OOBNs. Since weak conditional independence plays such an important role in OOBNs, in this paper we establish two useful results relating weak conditional independence with *weak multivalued dependency* in relational databases. The first result strengthens a previous result relating conditional independence and multivalued dependency. The second result takes a step towards showing that the complete axiomatization for weak multivalued dependency is also *complete* for full weak conditional independence.

1 Introduction

Recently, *object-oriented Bayesian networks* (OOBNs) [2] were introduced to facilitate the construction of large *Bayesian networks* [4, 8]. In this framework, each variable may be itself a Bayesian network, that is, OOBNs allow the value of a variable to be itself a distribution. To ensure that an OOBN defines a *unique* coarsened distribution, certain conditional independence assumptions were made in [2]. However, we will explicitly demonstrate here that conditional independence is *not* a necessary condition for ensuring consistency. *Weak conditional independence* was introduced in [6] and shown to be a necessary and sufficient condition for ensuring a unique coarsened distribution.

Since weak conditional independence plays such an important role in OOBNs, in this paper we establish two useful results relating weak conditional independence with *weak multivalued dependency* [1] in relational databases. The first result strengthens a previous result by Malvestuto [3] and Wong [5] relating conditional independence and multivalued dependency. The second result takes a step towards showing that the complete axiomatization for weak multivalued dependency [1] is also *complete* for full weak conditional independence.

This paper is organized as follows. Section 2 briefly outlines OOBNs and introduces the inconsistency problem. In Section 3, we review weak conditional independence. The relationship between weak conditional independence and weak multivalued dependency is studied in Section 4. The conclusion is presented in Section 5.

2 Object-Oriented Bayesian networks

Consider a finite set $U = \{A_1, A_2, \ldots, A_n\}$ of discrete random variables, where each variable $A \in U$ takes on values from a finite domain V_A. We may use capital letters, such as A, B, C, for variable names and lowercase letters a, b, c to denote specific values taken by those variables. Sets of variables will be denoted by capital letters such as X, Y, Z, and assignments of values to the variables in these sets (called configurations or tuples) will be denoted by lowercase letters x, y, z. We use V_X in the obvious way.

Let p be a *joint probability distribution* (jpd) [4] over the variables in U and X, Y, Z be subsets of U. We say Y and Z are *conditionally independent* given X, denoted $\mathbf{I}(Y, X, Z)$, if given any $x \in V_X$, $y \in V_Y$, then for all $z \in V_Z$,

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

A *Bayesian network* [4] is a *directed acyclic graph* (DAG) together with a corresponding set of conditional probability tables. The DAG graphically encodes conditional independencies assumed to hold in the problem domain. For example, based on the conditional independencies encoded in the Bayesian network in Fig. 1, the jpd $p(A, B, C, D, E)$ can be factorized as

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

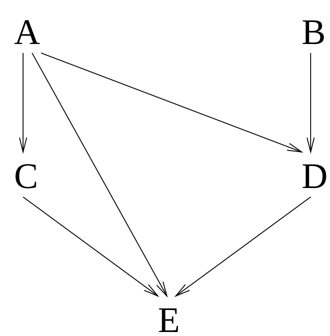

Fig. 1. A Bayesian network on variables $U = \{A, B, C, D, E\}$.

We now turn our attention to *granular* probabilistic networks [2]. We use the term granular to mean the ability to *coarsen* and *refine* parts of a probabilistic network.

Example 1. Consider the Bayesian network for a car accident [2] as shown in Fig. 2. One can *refine* the node *Car* to reveal the internal structure as shown in Fig. 3, where the rest of the network is not illustrated due to space limitations.

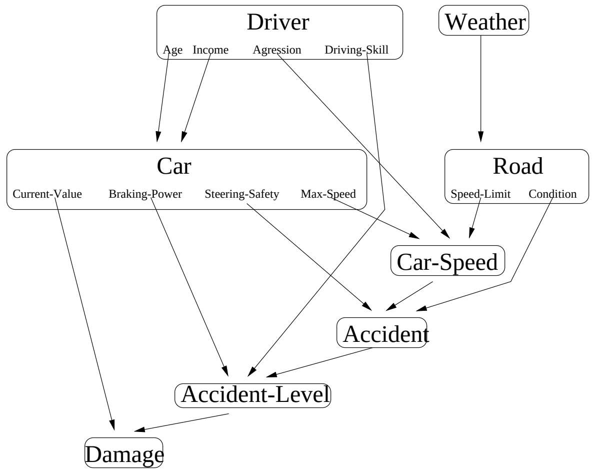

Fig. 2. A Bayesian network for a car accident.

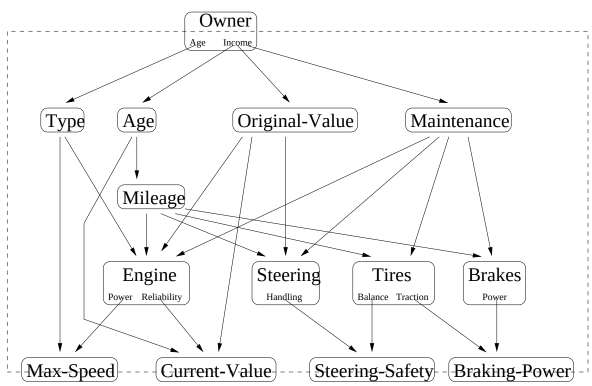

Fig. 3. Refining the Bayesian network in Fig. 2 to reveal the internal structure of the variable Car.

The *nest* operator [6], denoted ϕ, is used to *coarsen* a distribution $\mathbf{r}(XY)$. Intuitively, $\phi_{B=Y}(\mathbf{r})$ groups together the set Y of variables into a single variable B by forming a nested distribution given the same value for X.

Example 2. Consider $\mathbf{r}(ABC)$ in Fig. 4 (left). Coarsened distributions $\phi_{D=A}(\mathbf{r})$ (middle) and $\phi_{E=C}(\phi_{D=A}(\mathbf{r}))$ (right).

The main problem is that the nest operator does not necessarily commute.

Example 3. Recall $\mathbf{r}(ABC)$ in Fig. 4. It can be verified that $\phi_{D=A}(\phi_{E=C}(\mathbf{r})) \neq \phi_{E=C}(\phi_{D=A}(\mathbf{r}))$.

A	B	C	p(ABC)
0	0	0	0.4
1	0	0	0.4
1	0	1	0.2

D	B	C	p(DBC)
A p(A) 0 0.5 1 0.5	0	0	0.8
A p(A) 1 1.0	0	1	0.2

D	B	E	p(DBE)
A p(A) 0 0.5 1 0.5	0	C p(C) 0 1.0	0.8
A p(A) 1 1.0		C p(C) 1 1.0	0.2

Fig. 4. A joint distribution $\mathbf{r}(ABC)$ (left). Coarsened distributions $\phi_{D=A}(\mathbf{r})$ (middle) and $\phi_{E=C}(\phi_{D=A}(\mathbf{r}))$ (right).

In order to ensure that the nest operator commutes, i.e.,

$$\phi_{A=Y}(\phi_{B=Z}(\mathbf{r})) = \phi_{B=Z}(\phi_{A=Y}(\mathbf{r})), \tag{2}$$

assumptions must be made on the original joint distribution $\mathbf{r}$. Koller and Pfeffer [2] established that if the distribution $\mathbf{r}(YXZ)$ satisfies the conditional independence $\mathbf{I}(Y, X, Z)$, then Equation (2) holds. In other words, if Y and Z are conditionally independent given X, then $\phi_{A=Y}(\phi_{B=Z}(\mathbf{r})) = \phi_{B=Z}(\phi_{A=Y}(\mathbf{r}))$. The next example, however, explicitly demonstrates that conditional independence is *not* a necessary condition to ensure a unique coarsened distribution.

Example 4. The joint probability distribution $\mathbf{r}(ABC)$ in Fig. 5 (left) does *not* satisfy $\mathbf{I}(A, B, C)$, i.e., variables A and C are *not* conditionally independent given B. Nevertheless, it is still the case that $\phi_{D=A}(\phi_{E=C}(\mathbf{r}))$ in Fig. 5 (middle) is equal to $\phi_{E=C}(\phi_{D=A}(\mathbf{r}))$ in Fig. 5 (right).

A	B	C	p(ABC)
0	0	0	0.225
0	0	1	0.375
1	0	0	0.075
1	0	1	0.125
2	0	2	0.200

D	B	E	p(BDE)
A p(A) 0 0.75 1 0.25	0	C p(C) 0 0.375 1 0.625	0.8
A p(A) 2 1.0	0	C p(C) 2 1.0	0.2

D	B	E	p(BDE)
A p(A) 0 0.75 1 0.25	0	C p(C) 0 0.375 1 0.625	0.8
A p(A) 2 1.0	0	C p(C) 2 1.0	0.2

Fig. 5. Distribution $\mathbf{r}(ABC)$ does not satisfy $\mathbf{I}(A, B, C)$, yet $\phi_{D=A}(\phi_{E=C}(\mathbf{r}))$ (middle) is equal to $\phi_{E=C}(\phi_{D=A}(\mathbf{r}))$ (right).

3 Weak Conditional Independence

As pointed out in [6], when introducing nested distributions one steps outside the realm of conditional independence and into the realm of *weak conditional independence*. In the following definition of weak conditional independence, $\circ$ denotes the usual *composition* of relations, while $\theta(X)$ denotes the *equivalence relation* (partition) of $\mathbf{r}(XYZ)$ induced by the set X of variables [6].

We say variables Y and Z are *weakly independent* given X in $\mathbf{r}(XYZ)$, denoted $\mathbf{W}(Y, X, Z)$, if the following two conditions hold: (i) $\theta(XY) \circ \theta(XZ) = \theta(XZ) \circ \theta(XY)$, and (ii) for every equivalence class π in $\theta(XY) \circ \theta(XZ)$, the distribution $\mathbf{r}_\pi(XYZ)$ satisfies $\mathbf{I}(Y, X, Z)$, where $\mathbf{r}_\pi(XYZ)$ is defined as: $\mathbf{r}_\pi(XYZ) = \{ t \mid t \in \mathbf{r} \text{ and } t \in \pi \}$. Condition (i) says that $\theta(XY) \circ \theta(XZ)$ is an *equivalence relation*. Condition (ii) says that each equivalence class satisfies $\mathbf{I}(Y, X, Z)$.

Example 5. Let us verify that $\mathbf{r}(ABC)$ in Fig. 5 satisfies $\mathbf{W}(A, B, C)$. By definition, $\theta(AB) = \{ [t_1, t_2], [t_3, t_4], [t_5] \}$ and $\theta(BC) = \{ [t_1, t_3], [t_2, t_4], [t_5] \}$. Condition (i) is satisfied since

$$\theta(AB) \circ \theta(BC) = \{ \pi_1 = [t_1, t_2, t_3, t_4],\ \pi_2 = [t_5] \} \ = \ \theta(BC) \circ \theta(AB).$$

It can be verified that the (horizontal) distribution $\mathbf{r}_{\pi_1}(ABC) = \{t_1, t_2, t_3, t_4\}$ satisfies $\mathbf{I}(A, B, C)$. Since $\mathbf{r}_{\pi_2}(ABC) = \{t_5\}$ also satisfies $\mathbf{I}(A, B, C)$, by definition, $\mathbf{r}(ABC)$ satisfies $\mathbf{W}(A, B, C)$.

Weak conditional independence is very important in OOBNs, since it has been shown [6] that a distribution $\mathbf{r}(XYZ)$ satisfies $\mathbf{W}(Y, X, Z)$ if and only if $\phi_{A=Y}(\phi_{B=Z}(\mathbf{r})) = \phi_{B=Z}(\phi_{A=Y}(\mathbf{r}))$.

4 Properties of Weak Conditional Independence

In this section, we establish two properties of weak conditional independence in its relationship to *weak multivalued dependency* [1] in relational databases.

Before discussing weak conditional independence $\mathbf{W}(Y, X, Z)$ and weak multivalued dependency $W(Y, X, Z)$, let us first recall one result by Malvestuto [3] and Wong [5] relating conditional independence $\mathbf{I}(Y, X, Z)$ with *multivalued dependency* $I(Y, X, Z)$ in relational databases.

Theorem 1. [3, 5] Let $r(XYZ)$ be a relation with n tuples. Let $\mathbf{r}(XYZ)$ be the *uniform* joint distribution corresponding to $r(XYZ)$, i.e., the probability value of each tuple in $\mathbf{r}(XYZ)$ is $1/n$. Then $r(XYZ)$ satisfies the multivalued dependency $I(Y, X, Z)$ iff $\mathbf{r}(XYZ)$ satisfies the conditional independence $\mathbf{I}(Y, X, Z)$.

Since conditional independence is a special case of weak conditional independence and multivalued dependency is a special case of weak multivalued dependency, Theorem 1 is strengthened as follows.

Theorem 2. Let $r(XYZ)$ be a relation with n tuples. Let $\mathbf{r}(XYZ)$ be the *uniform* joint distribution corresponding to $r(XYZ)$, i.e., the probability value of each tuple in $\mathbf{r}(XYZ)$ is $1/n$. Then $r(XYZ)$ satisfies the weak multivalued dependency $W(Y, X, Z)$ iff $\mathbf{r}(XYZ)$ satisfies the weak conditional independence $\mathbf{W}(Y, X, Z)$.

Proof: ($\Rightarrow$) Suppose relation $r(XYZ)$ satisfies $W(Y, X, Z)$. From an equivalent definition of weak multivalued dependency [7], $\theta(XY) \circ \theta(XZ) = \theta(XZ) \circ \theta(XY)$. This implies that every equivalence class $\pi = \{t_1, t_2, \ldots, t_k\}$ in the equivalence relation $\theta(XY) \circ \theta(XZ)$ satisfies $I(Y, X, Z)$. Since $\mathbf{r}(XYZ)$ is a uniform distribution, the probabilistic distribution $\mathbf{r}_\pi(XYZ) = \{\ t \mid t \in \mathbf{r}(XYZ) \text{ and } t \in \pi\ \}$, is also a *constant* distribution. Since $\pi = \{t_1, t_2, \ldots, t_k\}$ satisfies $I(Y, X, Z)$, by Theorem 1, the constant distribution $\mathbf{r}_\pi(XYZ)$ satisfies $\mathbf{I}(Y, X, Z)$. This observation holds for every equivalence class in $\theta(XY) \circ \theta(XZ)$. Since $\theta(XY) \circ \theta(XZ) = \theta(XZ) \circ \theta(XY)$ and every equivalence class satisfies $\mathbf{I}(Y, X, Z)$, by definition, the uniform distribution $\mathbf{r}(XYZ)$ satisfies $\mathbf{W}(Y, X, Z)$.

($\Leftarrow$) Suppose the uniform distribution $\mathbf{r}(XYZ)$ satisfies $\mathbf{W}(Y, X, Z)$. By definition, $\theta(XY) \circ \theta(XZ) = \theta(XZ) \circ \theta(XY)$ and every equivalence class in $\theta(XY) \circ \theta(XZ)$ satisfies $\mathbf{I}(Y, X, Z)$. By the former, $r(XYZ)$ satisfies $W(Y, X, Z)$. $\square$

Corollary 3. In *non-uniform* distributions, weak multivalued dependency is a necessary but not a sufficient condition for weak conditional independence.

Theorem 2 says that a traditional relation $r(XYZ)$ satisfies $W(Y, X, Z)$ if and only if the uniform probabilistic distribution $\mathbf{r}(XYZ)$ satisfies $\mathbf{W}(Y, X, Z)$.

Example 6. The traditional relation $r(ABC)$ in Fig. 6 satisfies the weak multivalued dependency $W(A, B, C)$. By Theorem 2, the corresponding *uniform* distribution $\mathbf{r}(ABC)$ must satisfy the weak conditional independence $\mathbf{W}(A, B, C)$.

A	B	C
0	0	0
0	0	1
1	0	0
1	0	1
2	0	2

A	B	C	$p(ABC)$
0	0	0	0.2
0	0	1	0.2
1	0	0	0.2
1	0	1	0.2
2	0	2	0.2

Fig. 6. A traditional relation $r(ABC)$ and the corresponding the *uniform* distribution.

Our last objective in this exposition strives toward a complete axiomatization for the class of *full* weak conditional independence. Given $R = ABCD$, the weak independence $\mathbf{W}(A, B, CD)$ is called *full*, while $\mathbf{W}(A, B, C)$ is *embedded* since $ABC \subset R$. One approach would be to develop a set of axioms and establish completeness. Similar to [9], the approach we take here is to draw a one-to-one correspondence between weak conditional independence and weak multivalued dependency in relational databases.

The class of weak multivalued dependency has a *complete* axiomatization [1] consisting of four inference axioms (W1)-(W4). If we could establish that the *logical implication* [9], denoted $\models$, of (the numeric) weak conditional independence is exactly the same as that of (the non-numeric) weak multivalued dependency, then the complete axiomatization for weak multivalued dependency would also *complete* for full weak conditional independence. Hence, let the corresponding inferences axioms $(\mathbf{W1}) - (\mathbf{W4})$ for weak conditional independence be defined by replacing each occurrence of $W(Y, X, Z)$ in (W1)-(W4) by $\mathbf{W}(Y, X, Z)$.

Theorem 4. Given the *complete* axiomatization (W1)-(W4) for the weak multivalued dependency class. If the corresponding axioms $(\mathbf{W1})$-$(\mathbf{W4})$ for the weak conditional independence class are *sound* [9], then

$$\mathbf{C} \models \mathbf{c} \Longleftrightarrow C \models c,$$

where $\mathbf{C}$ is a set of weak conditional independencies, the corresponding set of weak multivalued dependencies is $C = \{W(Y, X, Z) \mid \mathbf{W}(Y, X, Z) \in \mathbf{C}\}$, and c is the weak multivalued dependency corresponding to a full weak conditional independence $\mathbf{c}$.

Proof: $(\Rightarrow)$ Suppose $\mathbf{C} \models \mathbf{c}$. We will prove the claim by contradiction. That is, suppose that $C \not\models c$. By definition, there exists a relation $r(R)$ such that $r(R)$ satisfies all of the weak multivalued dependencies in C, but $r(R)$ does not satisfy the weak multivalued dependency c. Let k denote the number of tuples in $r(R)$. We construct a distribution $\mathbf{r}(R)$ from $r(R)$ by appending the attribute A_p. For each of the k tuples in $\mathbf{r}(R)$, set $\mathbf{t}(A_p) = 1/k$. Thus, $\mathbf{r}(R)$ represents a *uniform* distribution. In the uniform case, by Theorem 2, $\mathbf{r}(R)$ satisfies $\mathbf{C}$ if and only if $r(R)$ satisfies C. Again using the uniform case, $\mathbf{r}(R)$ does not satisfy $\mathbf{c}$ since $r(R)$ does not satisfy c. By definition, $\mathbf{C}$ does not logically imply $\mathbf{c}$, namely, $\mathbf{C} \not\models \mathbf{c}$. A contradiction to the initial assumption that $\mathbf{C} \models \mathbf{c}$. Therefore, $C \models c$.
$(\Leftarrow)$ Let $C \models c$. Since (W1)-(W4) are *complete*, $C \models c$ implies that there exists a derivation sequence s of the weak multivalued dependency c by applying the weak multivalued dependency axioms to the weak multivalued dependencies in C. That c can be derived from C is written $C \vdash c$. On the other hand, each weak multivalued dependency axiom has a corresponding weak conditional independence axiom. This means there exists a derivation sequence $\mathbf{s}$ of the weak conditional independence $\mathbf{c}$ using the weak conditional independence axioms $(\mathbf{W1})$-$(\mathbf{W4})$ on the weak conditional independencies in $\mathbf{C}$, which parallels the derivation sequence s of the weak multivalued dependency c. That is, $\mathbf{C} \vdash \mathbf{c}$. Since the weak conditional independence axioms are assumed to be sound, $\mathbf{C} \vdash \mathbf{c}$ implies that $\mathbf{C} \models \mathbf{c}$. $\square$

5 Conclusion

Since *weak conditional independence* is a necessary and sufficient condition for the nest operator to commute [6], we established useful results relating weak conditional independence to weak multivalued dependency in relational databases. Theorem 2 strengthens a previous result by Malvestuto [3] and Wong [5] relating conditional independence and multivalued dependency. The other result (Theorem 4) shows that the complete axiomatization for weak multivalued dependency [1] is also *complete* for full weak conditional independence, provided that the corresponding weak conditional independence inference axioms are sound.

References

1. Fischer, P., Van Gucht, D.: Weak multivalued dependencies, *Proceedings of the Third ACM SIGACT-SIGMOD Symposium on the Principles of Database Systems*, 266–274, 1984.
2. Koller, D.,Pfeffer, A.: Object-oriented Bayesian networks. In *Thirteenth Conference on Uncertainty in Artificial Intelligence*, 302–313, 1997.
3. Malvestuto, F.: A unique formal system for binary decompositions of database relations, probability distributions and graphs. *Information Sciences*, 59:21–52, 1992.
4. Pearl, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, 1988.
5. Wong, S.K.M.: An extended relational data model for probabilistic reasoning. *Journal of Intelligent Information Systems*, 9:181–202, 1997.
6. Wong, S.K.M., Butz, C.J.: Contextual weak independence in Bayesian networks. In *Fifteenth Conference on Uncertainty in Artificial Intelligence*, 670–679, 1999.
7. Wong, S.K.M., Butz, C.J.: A comparative study of noncontextual and contextual dependencies. In *12th International Symposium on Methodologies for Intelligent Systems*, 247-255, 2000.
8. Wong, S.K.M., Butz, C.J.: Constructing the dependency structure of a multi-agent probabilistic network. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 13, No. 3, 395–415, 2001.
9. 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.