A Local Nest Property in Granular Probabilistic Networks

C.J. Butz and **S.K.M. Wong**
Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2, E-mail: {butz,wong}@cs.uregina.ca

Abstract: *In our earlier paper, we introduced the notion of granular probabilistic networks. We use the term granular to mean the ability to coarsen and refine parts of a joint probability distribution. In practice, however, the joint distribution is usually represented as a product of marginal distributions. In this paper, we show that the coarsening operation can be applied locally to the marginal distributions with the same effect as if applied to the joint distribution. This is an important result as otherwise the joint distribution would have to be computed before performing the coarsening operation.*

1 Introduction

Probabilistic networks [2, 5] have become a proven tool for uncertainty management. Since it is not always feasible to store a joint probability distribution explicitly, the notion of conditional independence is utilized to express a joint probability distribution as a product of several smaller distributions. A *Bayesian network* [5] represents a joint distribution using conditional probability distributions, while a *Markov* network [2] represents a joint distribution using marginal distributions. Recently, attempts have been made to generalize Bayesian networks using nested distributions [3, 7].

In our earlier paper [7], we introduced the notion of *granular* probabilistic networks. We use the term granular to mean the ability to *coarsen* and *refine* parts of a joint probability distribution. These operations were defined with respect to the joint distribution. In this paper, we show that the coarsening operation can be applied *locally* to the marginal distributions with the same effect as if applied to the joint distribution itself. This is an important result as otherwise the joint distribution would have to be computed before performing the coarsening operation.

This paper is organized as follows. In Section 2, we review our Bayesian database model, which unifies relational databases and Bayesian networks. Granular probabilistic networks are introduced in Section 3. In Section 4, we present the main result of this paper. The conclusion is given in Section 5.

2 A Bayesian Database Model

Our *Bayesian database model* [6, 8] provides a unified framework for modeling both relational database and probabilistic reasoning applications. We extend the standard relation by adding a column to store the number of duplicate tuples (probability values). The generalized relation represents a joint probability distribution in Bayesian networks. The relational database operators are generalized to take the probability values into consideration. In the relational database model, the lossless decomposition of a relation into two projections (smaller relations) is based upon the notion of *embedded multivalued dependency* (EMVD) [4]. In our model, the *Bayesian embedded multivalued dependency* (BEMVD) is the necessary and sufficient condition for the lossless decomposition of a generalized relation into two smaller generalized relations. It is important to note that BEMVD is simply probabilistic conditional independence expressed as a database dependency.

We say a generalized relation $\mathbf{r}(XYZW)$ satisfies the *Bayesian embedded multivalued dependency* (BEMVD), $X \Rightarrow\Rightarrow Y|Z$, if

$$\begin{aligned} \tau_{XYZ}(\mathbf{r}) &= \tau_{XY}(\mathbf{r}) \otimes \tau_{XZ}(\mathbf{r}) \\ &\equiv \tau_{XY}(\mathbf{r}) \times \tau_{XZ}(\mathbf{r}) \times \tau_{X}(\mathbf{r})^{-1}, \end{aligned}$$

where τ and $\times$ are the *marginalization* and *product join* operators, respectively. For example, the relation $\mathbf{r}(A_1A_2A_3)$ in Figure 1 represents a probability distribution $p(A_1A_2A_3)$. Furthermore, relation $\mathbf{r}(A_1A_2A_3)$ satisfies the BEMVD $A_2 \Rightarrow\Rightarrow A_1|A_3$, since $\mathbf{r}(A_1A_2A_3)$ can be decomposed as

$$\mathbf{r}(A_1A_2A_3) \quad = \quad \tau_{A_1A_2}(\mathbf{r}) \otimes \tau_{A_2A_3}(\mathbf{r}),$$

as shown in Figure 2.

Let $\mathcal{R} = \{R_1, R_2, \ldots, R_n\}$ be an *acyclic hypergraph* [1, 2] on the set of attributes $R = R_1 \cup R_2 \cup \ldots \cup R_n$. We say a *Bayesian acyclic join dependency* (BAJD), $\otimes\mathcal{R}$, is satisfied by a relation $\mathbf{r}(R)$, if

$$\mathbf{r}(R) = (\ldots((\tau_{R_1}(\mathbf{r}) \otimes \tau_{R_2}(\mathbf{r})) \otimes \tau_{R_3}(\mathbf{r}))\ldots) \otimes \tau_{R_n}(\mathbf{r}),$$

where the sequence $R_1, R_2, \ldots, R_n$ is a hypertree construction ordering for $\mathcal{R}$. (We will omit the parentheses if a hypertree construction ordering is understood.) For example, relation $\mathbf{r}(A_1A_2A_3A_4A_5A_6)$ in

A_1	A_2	A_3	A_p
0	0	0	0.375
0	0	1	0.125
1	0	0	0.225
1	0	1	0.075
1	1	0	0.200

Figure 1: A relation $\mathbf{r}(A_1A_2A_3)$.

A_1	A_2	$A_{p(R_1)}$
0	0	0.5
1	0	0.3
1	1	0.2

$\otimes$

A_2	A_3	$A_{p(R_2)}$
0	0	0.6
0	1	0.2
1	0	0.2

Figure 2: Relation $\mathbf{r}(A_1A_2A_3)$ in Figure 1 satisfies the BEMVD $A_2 \Rightarrow\Rightarrow A_1|A_3$.

Figure 3 satisfies the BAJD $\otimes\mathcal{R}$ since

$$\mathbf{r}(R) \quad = \quad \tau_{R_1}(\mathbf{r}) \otimes \tau_{R_2}(\mathbf{r}) \otimes \tau_{R_3}(\mathbf{r}) \otimes \tau_{R_4}(\mathbf{r}),$$

where $\mathcal{R} = \{R_1 = A_1A_2A_3, R_2 = A_1A_2A_4, R_3 = A_1A_3A_5, R_4 = A_2A_3A_6\}$ is an acyclic hypergraph on the set $R = A_1A_2A_3A_4A_5A_6$ of attributes.

3 Granular Probabilistic Networks

In this section, we present a brief discussion on the recent attempts to generalize Bayesian networks with *complex-values* (nested distributions) [3, 7]. That is, the value of an attribute is allowed to be another distribution itself. We introduce operators called *nest* and *unnest* to *coarsen* and *refine* parts of a probabilistic network, respectively.

Koller and Pfeffer [3] suggested a framework for the modeling of and inference in a large Bayesian network. In this framework, parts of the network can be coarsened and refined. For example, consider the Bayesian network for a car accident [3] as shown in Figure 4. One can *refine* the node *Car* to reveal the internal structure as shown in Figure 5, where the rest of the network is not shown due to space limitations.

$\mathbf{r}(R) =$

A_1	A_2	A_3	A_4	A_5	A_6	A_p
0	0	0	0	0	0	0.1
0	0	1	0	1	1	0.2
0	1	0	1	0	2	0.3
1	0	0	2	2	0	0.4

Figure 3: Relation $\mathbf{r}(R)$ satisfies the BAJD $\otimes\{A_1A_2A_3, A_1A_2A_4, A_1A_3A_5, A_2A_3A_6\}$.

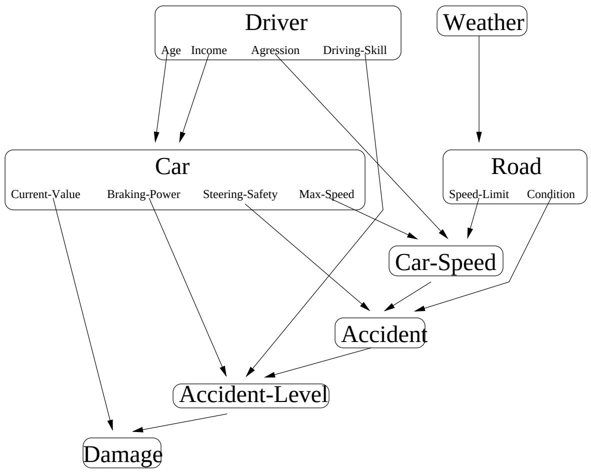

Figure 4: A Bayesian network for a car accident.

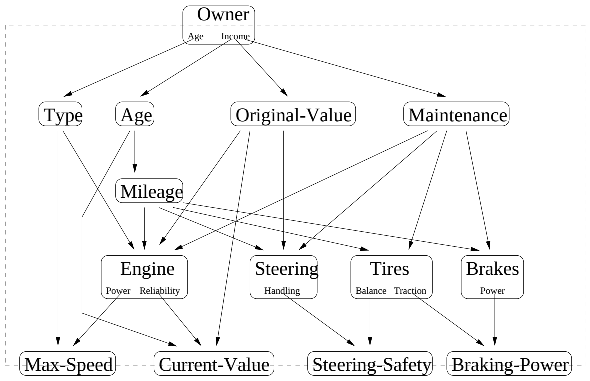

Figure 5: Refining the Bayesian network in Figure 4 to reveal the internal structure of the variable *Car*.

These ideas were formalized in our Bayesian database model [7]. Here we review two operators, called *nest* and *unnest*, for coarsening and refining a probabilistic network, respectively.

The *nest* operator ϕ is used to *coarsen* a relation $\mathbf{r}(XY)$. Intuitively, $\phi_{B=Y}(\mathbf{r})$ groups together all the Y-values into a nested relation given the same X-value.

Let $\mathbf{r}(XY)$ be a relation with X and Y disjoint sets of attributes. Coarsening attributes Y in $\mathbf{r}(XY)$ as attribute B is the relation $\phi_{B=Y}(\mathbf{r})$ with attributes XB defined as

$$\phi_{B=Y}(\mathbf{r}) \quad = \quad \{t \mid t(X) = u(X) \text{ and } t(B) = \{u(YA_p)\} \text{ and } t(A_{p(XB)}) = \sum_u u(A_p)\},$$

where $u \in \mathbf{r}$. The attribute A_p in the B-value is relabelled $A_{p(Y)}$ and the values normalized. For example,

$\phi_{B=A_3}(\mathbf{r})$ is shown in Figure 6, where $\mathbf{r}(A_1A_2A_3)$ is the relation in Figure 1.

<table>
<tr><th>A_1</th><th>A_2</th><th>B</th><th>$A_{p(A_1A_2B)}$</th></tr>
<tr><td>0</td><td>0</td><td><table><tr><th>A_3</th><th>$A_{p(A_3)}$</th></tr><tr><td>0</td><td>0.75</td></tr><tr><td>1</td><td>0.25</td></tr></table></td><td>0.5</td></tr>
<tr><td>1</td><td>0</td><td><table><tr><th>A_3</th><th>$A_{p(A_3)}$</th></tr><tr><td>0</td><td>0.75</td></tr><tr><td>1</td><td>0.25</td></tr></table></td><td>0.3</td></tr>
<tr><td>1</td><td>1</td><td><table><tr><th>A_3</th><th>$A_{p(A_3)}$</th></tr><tr><td>0</td><td>1.0</td></tr></table></td><td>0.2</td></tr>
</table>

Figure 6: The relation $\phi_{B=A_3}(\mathbf{r})$.

Whereas nest will coarsen a relation, the *unnest* operator ψ *refines* a relation. Intuitively, $\psi_{B=Y}(\mathbf{r})$ flattens relation $\mathbf{r}(XB)$ by joining each X-value with each Y-value in the nested B-value. Revealing the nested variables Y in attribute B of relation $\mathbf{r}(XB)$ is the relation $\psi_{B=Y}(\mathbf{r})$ with attributes XY defined as

$$\begin{aligned}\psi_{B=Y}(\mathbf{r}) &= \{t \mid t(X) = u(X) \text{ and } t(Y) \in u(B) \\ &\quad \text{and } t(A_p) = v(A_{p(Y)}) \cdot u(A_p)\},\end{aligned}$$

where $u \in \mathbf{r}$, $v \in u(B)$, and $v(Y) = t(Y)$. Note that we may write $\psi_{B=Y}(\mathbf{r})$ as $\psi_B(\mathbf{r})$ since Y is implied by B. For example, revealing the nested attribute A_3 in relation $\mathbf{r}(A_1A_2B)$ in Figure 6 gives the relation $\psi_B(\mathbf{r}(A_1A_2B)) = \mathbf{r}(A_1A_2A_3)$ in Figure 1.

4 Nesting Locally

The results presented thus far have been defined with respect to the joint probability distribution $\mathbf{r}(R)$. In practice, however, $\mathbf{r}(R)$ is represented as a BAJD [8]. That is, the joint probability distribution is factorized in terms of marginal distributions over an acyclic hypergraph. The following main result of this paper indicates that the nest operation can be applied *locally* to marginal distributions with the same result as if applied to $\mathbf{r}(R)$ itself.

Theorem 1 Let $\mathbf{r}(R)$ satisfy a BAJD $\otimes\mathcal{R} = \{R_1, R_2, \ldots, R_n\}$. Let $Z \subseteq R_n$ such that $Z \cap R_j = \emptyset$, $j = 1, 2, \ldots, n-1$. Then

$$\begin{aligned}&\phi_{B=Z}(\mathbf{r}) \\ =\ &\tau_{R_1}(\mathbf{r}) \otimes \tau_{R_2}(\mathbf{r}) \otimes \ldots \otimes \tau_{R_{n-1}}(\mathbf{r}) \otimes \phi_{B=Z}(\tau_{R_n}(\mathbf{r})).\end{aligned}$$

Proof: Suppose $\mathbf{r}(R)$ satisfies a BAJD $\otimes\mathcal{R}$. Without loss of generality, let $\mathbf{r}(XYZ) = \tau_{XY}(\mathbf{r}) \otimes \tau_{XZ}(\mathbf{r})$, where $\mathbf{r}(XYZ)$ is shown in Figure 7 and $\tau_{XY}(\mathbf{r}) \otimes \tau_{XZ}(\mathbf{r})$ is shown in Figure 8. Since the BAJD is satisfied, we have $c = a_1 + a_2 = b_1 + b_2$. Computing $\phi_{B=Z}(\mathbf{r})$ using the joint distribution $\mathbf{r}$ is shown in Figure 9. On the other hand, $\tau_{XY}(\mathbf{r}) \otimes \phi_{B=Z}(\tau_{XZ}(\mathbf{r}))$, where $\phi_{B=Z}(\tau_{XZ}(\mathbf{r}))$ is computed using the marginal distribution $\tau_{XZ}(\mathbf{r})$, is illustrated in Figure 10. We obtain our desired result since,

$$\phi_{B=Z}(\mathbf{r}) \quad = \quad \tau_{XY}(\mathbf{r}) \otimes \phi_{B=Z}(\tau_{XZ}(\mathbf{r})). \ \square$$

Y	X	Z	A_p
y_1	x_1	z_1	a_1b_1/c
y_1	x_1	z_2	a_1b_2/c
y_2	x_1	z_1	a_2b_1/c
y_2	x_1	z_2	a_2b_2/c

Figure 7: Relation $\mathbf{r}(XYZ)$ representing a joint distribution.

Y	X	$A_{p(YX)}$
y_1	x_1	b_1
y_2	x_1	b_2

$\otimes$

X	Z	$A_{p(XZ)}$
x_1	z_1	a_1
x_1	z_2	a_2

Figure 8: Relation $\mathbf{r}(YXZ)$ in Figure 7 satisfies the BAJD $\otimes\{YX, XZ\}$.

<table>
<tr><th>Y</th><th>X</th><th>B</th><th>$A_{p(YXB)}$</th></tr>
<tr><td>y_1</td><td>x_1</td><td><table><tr><th>Z</th><th>$A_{p(Z)}$</th></tr><tr><td>z_1</td><td>$a_1b_1/c = b_1$</td></tr><tr><td>z_2</td><td>$a_1b_2/c = b_2$</td></tr></table></td><td>a_1</td></tr>
<tr><td>y_2</td><td>x_1</td><td><table><tr><th>Z</th><th>$A_{p(Z)}$</th></tr><tr><td>z_1</td><td>$a_2b_1/c = b_1$</td></tr><tr><td>z_2</td><td>$a_2b_2/c = b_2$</td></tr></table></td><td>a_2</td></tr>
</table>

Figure 9: Relation $\phi_{B=Z}(\mathbf{r})$, where $\mathbf{r}(R)$ is the joint distribution in Figure 7.

This is an important result as otherwise the joint distribution would have to be computed before performing the nest operation. As in the traditional probabilistic model, it is not always feasible to store the joint probability distribution. Theorem 1 indicates that performing the nest operation locally on

<table>
<tr><th>Y</th><th>X</th><th>$A_{p(YX)}$</th></tr>
<tr><td>y_1</td><td>x_1</td><td>a_1</td></tr>
<tr><td>y_2</td><td>x_1</td><td>a_2</td></tr>
</table>

$\otimes$

<table>
<tr><th>X</th><th>B</th><th>$A_{p(XB)}$</th></tr>
<tr><td>x_1</td><td><table>
<tr><th>Z</th><th>$A_{p(Z)}$</th></tr>
<tr><td>z_1</td><td>b_1</td></tr>
<tr><td>z_2</td><td>b_2</td></tr>
</table></td><td>c</td></tr>
</table>

Figure 10: The relation on the right is $\phi_{B=Z}(\tau_{XZ}(\mathbf{r}))$, where $\tau_{XZ}(\mathbf{r})$ is a marginal distribution.

the factorized representation of the joint distribution is equivalent to performing the nest operation on the joint distribution itself.

We illustrate Theorem 1 with an example. Recall the generalized relation $\mathbf{r}(A_1A_2A_3)$ in Figure 1. Relation $\mathbf{r}(A_1A_2A_3)$ satisfies the BEMVD $A_2 \Rightarrow\Rightarrow A_1|A_3$, since $\mathbf{r}(A_1A_2A_3)$ can be decomposed as

$$\mathbf{r}(A_1A_2A_3) \quad = \quad \tau_{A_1A_2}(\mathbf{r}) \otimes \tau_{A_2A_3}(\mathbf{r}),$$

as shown in Figure 2. Nesting attribute A_3 locally in the marginal distribution $\tau_{A_2A_3}(\mathbf{r})$ is shown on the right of Figure 11. It can be easily verified that computing $\tau_{A_1A_2}(\mathbf{r}) \otimes \phi_{B=A_3}(\tau_{A_2A_3}(\mathbf{r}))$ gives the same relation as shown in Figure 6.

<table>
<tr><th>A_1</th><th>A_2</th><th>$A_{p(A_1A_2)}$</th></tr>
<tr><td>0</td><td>0</td><td>0.5</td></tr>
<tr><td>1</td><td>0</td><td>0.3</td></tr>
<tr><td>1</td><td>1</td><td>0.2</td></tr>
</table>

$\otimes$

<table>
<tr><th>A_2</th><th>B</th><th>$A_{p(A_2B)}$</th></tr>
<tr><td>0</td><td><table>
<tr><th>A_3</th><th>$A_{p(A_3)}$</th></tr>
<tr><td>0</td><td>0.75</td></tr>
<tr><td>1</td><td>0.25</td></tr>
</table></td><td>0.8</td></tr>
<tr><td>1</td><td><table>
<tr><th>A_3</th><th>$A_{p(A_3)}$</th></tr>
<tr><td>0</td><td>1.0</td></tr>
</table></td><td>0.2</td></tr>
</table>

Figure 11: $\tau_{A_1A_2}(\mathbf{r}) \otimes \phi_{B=A_3}(\tau_{A_2A_3}(\mathbf{r})) = \phi_{B=A_3}(\mathbf{r})$, where $\phi_{B=A_3}(\mathbf{r})$ is shown in Figure 6.

5 Conclusion

Granular probabilistic networks have recently been proposed [3, 7]. That is, parts of a joint distribution can be *coarsened* or *refined* depending on whether they are of current interest or not. Since it is not always feasible to store a joint probability distribution explicitly, in practice a joint distribution is factorized into smaller distributions. The main result of this paper is that the coarsening operation can be applied *locally* to the marginal distributions with the same effect as if applied to the joint distribution itself. This is an important result as otherwise the joint distribution would have to be computed before performing the coarsening operation.

References

[1] C. Beeri, R. Fagin, D. Maier, and M. Yannakakis. On the desirability of acyclic database schemes. *Journal of the ACM*, 30(3):479–513, July 1983.

[2] P. Hajek, T. Havranek, and R. Jirousek. *Uncertain Information Processing in Expert Systems.* CRC Press, 1992.

[3] D. Koller and A. Pfeffer. Object-oriented bayesian networks. In *Thirteenth Conference on Uncertainty in Artificial Intelligence*, pages 302–313. Morgan Kaufmann Publishers, 1997.

[4] D. Maier. *The Theory of Relational Databases.* Principles of Computer Science. Computer Science Press, Rockville, Maryland, 1983.

[5] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, San Francisco, California, 1988.

[6] S.K.M. Wong. An extended relational data model for probabilistic reasoning. *Journal of Intelligent Information Systems*, 9:181–202, 1997.

[7] S.K.M. Wong and C.J. Butz. Contextual weak independence in bayesian networks. In *Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 670–679. Morgan Kaufmann Publishers, 1999.

[8] S.K.M. Wong, C.J. Butz, and Y. Xiang. A method for implementing a probabilistic model as a relational database. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 556–564. Morgan Kaufmann Publishers, 1995.