A General Coarsening Method for Granular Probabilistic Networks

C.J. Butz
Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2, E-mail: butz@cs.uregina.ca

Abstract *In this paper, we present a general method for coarsening a set of variables in a probabilistic network. The work here is an extension of our earlier works, which always placed restrictions on the sets to be coarsened. The soundness of our method is also shown.*

1 Introduction

Due to their rigorous mathematical foundation, *Bayesian networks* [4, 7, 8] have gained wide acceptance in practice for uncertainty management. In [6], we coined the phrase *granular* probabilistic reasoning to mean the ability to *coarsen* and *refine* parts of a probabilistic network depending on whether they are of interest or not. Granular probabilistic reasoning is of the utmost importance as it facilitates the design of *large* BNs [3]. In addition, granular probabilistic reasoning may lead to more efficient probabilistic inference [3]. It is then not surprising that Xiang, in his recent text [9], explicitly states that granular probabilistic reasoning [3, 6] demands further attention.

In [6], we proposed two operators called *nest* and *unnest* for coarsening and refining variables in a network, respectively. We showed in [1] that the nest operator can be applied *locally* to a marginal distribution with the same effect as if it were applied directly to the joint distribution. A method for coarsening variables spread throughout a probabilistic network was presented in [2], albeit with restrictions imposed. To the best of our knowledge, no study has ever put forth a method to coarsen variables in a probabilistic network without imposing any conditions.

In this paper, we present a general method for coarsening variables in a probabilistic network. Unlike the ones suggested in [1, 2], the proposed method here does not impose any conditions on the set to be coarsened. This method gathers all variables to be coarsened into one marginal distribution, and then applies the nest operator. In Theorem 1, we show our method is correct.

This paper is organized as follows. Section 2 reviews the nest operator. Related works are discussed in Section 3. In Section 4, we present a general method for nesting variables. The conclusion is presented in Section 5.

2 The Nest Operator

Consider the joint distribution $p(R)$ represented as a probabilistic relation $\mathbf{r}(R)$ in Figure 1, where $R = \{A,B,C,D,E,F\} = ABCDEF$ is a set of variables. Variables A, B and C each have domain $\{0,1\}$, while D, E and F each have domain $\{0,1,2\}$. Configurations with zero probability are not shown.

A	B	C	D	E	F	$p(R)$
0	0	0	0	0	0	0.05
0	0	0	0	1	0	0.05
0	0	1	0	1	1	0.20
0	1	0	1	0	2	0.15
0	1	0	1	1	2	0.15
1	0	0	2	2	0	0.40

Figure 1. A probabilistic relation $\mathbf{r}(R)$ representing a joint distribution $p(R)$.

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

$$\phi_{A=Y}(\mathbf{r}) = \{t \mid t(X) = u(X),\ t(A) = \{u(Yp(R))\}, \quad t(p(XA)) = \sum_u u(p(R)),\ \text{and } u \in \mathbf{r}\}.$$

Attribute $p(R)$ in the A-value is relabeled $p(Y)$ and the values are normalized.

For example, consider the relation $\mathbf{r}(ABCDEF)$ in Figure 1. Nesting the set $Y = ADE$ of variables as the single variable G is the nested relation $\phi_{G=\{A,D,E\}}(\mathbf{r})$ in Figure 2. For instance, given the X-value $< B : 0, C : 0; F : 0 >$, the Y-values $< A : 0, D : 0, E : 0, p(R) : 0.05 >$ $< A : 0, D : 0, E : 1, p(R) : 0.05 >$ $< A : 1, D : 2, E : 2, p(R) : 0.40 >$ are grouped into a nested distribution. Here the attribute $p(R)$ is relabeled as $p(ADE)$, with the probability values 0.05, 0.05, 0.40 normalized as 0.10, 0.10, 0.80.

<table>
<tr><th>B</th><th>C</th><th>G</th><th>F</th><th>$p(W)$</th></tr>
<tr><td>0</td><td>0</td><td><table><tr><th>A</th><th>D</th><th>E</th><th>$p(X)$</th></tr><tr><td>0</td><td>0</td><td>0</td><td>0.1</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0.1</td></tr><tr><td>1</td><td>2</td><td>2</td><td>0.8</td></tr></table></td><td>0</td><td>0.5</td></tr>
<tr><td>0</td><td>1</td><td><table><tr><th>A</th><th>D</th><th>E</th><th>$p(X)$</th></tr><tr><td>0</td><td>0</td><td>1</td><td>1.0</td></tr></table></td><td>1</td><td>0.2</td></tr>
<tr><td>1</td><td>0</td><td><table><tr><th>A</th><th>D</th><th>E</th><th>$p(DE)$</th></tr><tr><td>0</td><td>1</td><td>0</td><td>0.5</td></tr><tr><td>0</td><td>1</td><td>1</td><td>0.5</td></tr></table></td><td>0</td><td>0.3</td></tr>
</table>

Figure 2. Relation $\phi_{G=\{A,D,E\}}(\mathbf{r})$, where $\mathbf{r}$ is in Figure 1, $X = ADE$, and $W = BCGF$.

3 Earlier Works

In practice, the joint distribution $\mathbf{r}(R)$ in Figure 1 is factorized using *conditional independencies* [7]. We say Y and Z are *conditionally independent* given X in the joint distribution $p(YXZ)$, denoted $I(Y, X, Z)$, if

$$p(YXZ) \quad = \quad \frac{p(XY) \cdot p(YZ)}{p(Y)}. \qquad (1)$$

For example, the joint distribution $p(R)$ in Figure 1 satisfies the conditional independence $I(D, AB, CEF)$ as

$$p(R) \quad = \quad \frac{p(ABD) \cdot p(ABCEF)}{p(AB)}.$$

In other words, variables D and CEF are conditionally independent given AB. Since $p(R)$ can also be factorized as

$$p(R) \quad = \quad \frac{p(ACE) \cdot p(ABCDF)}{p(AC)}$$

and

$$p(R) \quad = \quad \frac{p(ABCDE) \cdot p(BCF)}{p(BC)},$$

the joint distribution $p(R)$ also satisfies the conditional independencies $I(E, AC, BDF)$ and $I(ADE, BC, F)$.

The conditional independencies satisfied by $\mathbf{r}(R)$ can be graphically represented by an *acyclic hypergraph* [8]. For example, $\mathcal{R} = \{R_1 = \{A, B, D\}, R_2 = \{A, B, C\}, R_3 = \{A, C, E\}, R_4 = \{B, C, F\}\}$ is an acyclic hypergraph on $R = ABCDEF$. The *marginal* [7, 8] distributions $\mathbf{r}_1(ABD)$, $\mathbf{r}_2(ABC)$, $\mathbf{r}_3(ACE)$, and $\mathbf{r}_4(BCF)$ of $\mathbf{r}(R)$ are shown in Figure 3. The joint distribution $p(R)$ is then expressed as

$$p(R) = \frac{p(ABD) \cdot p(ABC) \cdot p(ACE) \cdot p(BCF)}{p(AB) \cdot p(AC) \cdot p(BC)},$$

which is an example of a *Markov network* [4]. In our probabilistic relational model [8], this Markov network is expressed as

$$\mathbf{r}(R) = \\ ((\mathbf{r}_1(ABD) \otimes \mathbf{r}_2(ABC)) \otimes \mathbf{r}_3(ACE)) \otimes \mathbf{r}_4(BCF),$$

where the Markov join operator $\otimes$ means

$$\mathbf{r}(XY) \otimes \mathbf{r}(YZ) \quad = \quad \frac{p(XY) \cdot p(YZ)}{p(Y)}.$$

We omit the parentheses for simplicity. For example, the Markov join $\mathbf{r}_1(ABD) \otimes \mathbf{r}_2(ABC) \otimes \mathbf{r}_3(ACE)$ is shown in Figure 4. That is,

$$p(ABCDE) = \frac{p(ABD) \cdot p(ABC) \cdot p(ACE)}{p(AB) \cdot p(AC)}.$$

A	B	D	$p(R_1)$
0	0	0	0.3
0	1	1	0.3
1	0	2	0.4

B	C	F	$p(R_4)$
0	0	0	0.5
0	1	1	0.2
1	0	2	0.3

A	B	C	$p(R_2)$
0	0	0	0.1
0	0	1	0.2
0	1	0	0.3
1	0	0	0.4

A	C	E	$p(R_3)$
0	0	0	0.2
0	0	1	0.2
0	1	1	0.2
1	0	2	0.3

Figure 3. The marginals $\mathbf{r}_1(ABD)$, $\mathbf{r}_2(ABC)$, $\mathbf{r}_3(ACE)$, and $\mathbf{r}_4(BCF)$ of relation $\mathbf{r}$.

A	B	C	D	E	$p(ABCDE)$
0	0	0	0	0	0.05
0	0	0	0	1	0.05
0	0	1	0	1	0.20
0	1	0	1	0	0.15
0	1	0	1	1	0.15
1	0	0	2	2	0.40

Figure 4. The Markov join $\mathbf{r}_1(ABD) \otimes \mathbf{r}_2(ABC) \otimes \mathbf{r}_3(ACE)$.

In [1], we gave a method for coarsening a set X of variables provided that the following two conditions were satisfied: (i) X appeared together in one table, and (ii) every variable in X appeared in no other table. This work was then extended as follows. A method was given in [2] for coarsening a set X of variables provided that every variable in X appeared in precisely one table. For example, the set DE could be nested since variable D only appears in table $\mathbf{r}_1(ABD)$ while E only appears in $\mathbf{r}_3(ACE)$. On the contrary, the set ADE could not be nested as variable A appears in the three tables $\mathbf{r}_1(ABD)$ $\mathbf{r}_2(ABC)$ and $\mathbf{r}_3(ACE)$. In the next section, we present a novel method to coarsen any subset of variables.

4 General Nesting

Since the nest operator is unary, the first task is to combine the set Y of variables into a single table. The *selective reduction algorithm* (SRA) [5] in relational databases is applied for this purpose. The nest operator can then be applied to coarsen Y as attribute A.

The input $(\mathcal{R}, Y)$ to the SRA algorithm is an acyclic hypergraph $\mathcal{R} = \{R_1, R_2, \ldots, R_n\}$ on the set R of variables and a subset $Y \subseteq R$ of variables to be coarsened. First, mark every variable A in Y. Second, apply the following two operations until neither can be applied: (i) delete an *unmarked* variable A if A appears in precisely one node R_i, and (ii) delete node R_i if it is a subset of another node R_j, $i \neq j$.

For example, suppose we wish to selectively reduce the acyclic hypergraph $\mathcal{R} = \{R_1 = \{A, B, D\}, R_2 = \{A, B, C\}, R_3 = \{A, C, E\}, R_4 = \{B, C, F\}\}$ to the set $Y = ADE$, i.e., compute SRA($\mathcal{R}, ADE$). First, mark the variables A, D and E. By operation (i), variable F can be deleted as it only appears in R_4. Thus, $R_4 = BC$. By operation (ii), R_4 can be deleted as it is contained by $R_2 = ABC$. No other operations can be applied. Thus, the output is $\mathcal{R} = \{R_1 = \{A, B, D\}, R_2 = \{A, B, C\}, R_3 = \{A, C, E\}\}$.

In general, some variables in R_1, R_2 and R_3 may be deleted. However, for our purposes here, we are only interested in those R_i which are not deleted by operation (ii) of the selective reduction algorithm. The reason being those variables in the remaining R_i must be joined to get the variables to be coarsened into a single table.

We now present the formal algorithm, called *General Nest* (GN). Given a Markov network $\mathcal{R}$ on the set R of variables, the GN algorithm will coarsen $Y \subseteq R$ as variable A.

Algorithm 1 General Nest($\mathcal{R}, Y, A$)
1. Let $R_j, \ldots, R_n$ be those elements of $\mathcal{R}$ not deleted by the call SRA($Y, \mathcal{R}$).
2. Compute the Markov join $\mathbf{r}_{jn} = \mathbf{r}_j(R_j) \otimes \ldots \otimes \mathbf{r}_n(R_n)$.
3. Compute $\phi_{A=Y}(\mathbf{r}_{jn})$.

For example, suppose we wish to compute GN($\mathcal{R}, ADE, G$). As previously mentioned, the selective reduction of $\mathcal{R}$ to the set ADE leaves $\{ABD, ABC, ACE\}$. The Markov join $\mathbf{r}_1(ABD) \otimes \mathbf{r}_2(ABC) \otimes \mathbf{r}_3(ACE)$ is shown in Figure 4. The nested relation $\phi_{G=\{A,D,E\}}(\mathbf{r}_1 \otimes \mathbf{r}_2 \otimes \mathbf{r}_3)$ is depicted in Figure 5.

Theorem 1 Let $\mathbf{r}(R) = \mathbf{r}_1(R_1) \otimes \ldots \otimes \mathbf{r}_i(R_i) \otimes \mathbf{r}_j(R_j) \ldots \otimes \mathbf{r}_n(R_n)$ be represented as a Markov network, and let $Y \subseteq R$. Then $\phi_{A=Y}(\mathbf{r})$ is the same as $\mathbf{r}_1(R_1) \otimes \ldots \otimes \mathbf{r}_i(R_i) \otimes \phi_{A=Y}(\mathbf{r}_{jn}(R_{jn}))$, where $\mathbf{r}_{jn}(R_{jn})$ is the Markov join of the nodes of the acyclic hypergraph remaining in the selective reduction of Y in $\{R_1, \ldots, R_n\}$.

Proof: Without loss of generality, suppose the selective reduction on $\{R_1, \ldots, R_i, R_j, \ldots, R_n\}$ deletes nodes $R_1, \ldots, R_i$ and leaves nodes $R_j, \ldots, R_n$. By the conditional independencies holding in the acyclic hypergraph, the Markov join $\mathbf{r}_j(R_j) \otimes \ldots \otimes \mathbf{r}_n(R_n)$ is the *marginal* distribution $\mathbf{r}(R_j \cup \ldots \cup R_n)$. Thus, the original factorization of $\mathbf{r}(R)$ can be rewritten as $\mathbf{r}(R) = \mathbf{r}_1(R_1) \otimes \ldots \otimes \mathbf{r}_i(R_i) \otimes \mathbf{r}(R_j \cup \ldots \cup R_n)$. Since $Y \subseteq (R_j \cup \ldots \cup R_n)$ and $Y \cap R_k = \emptyset$, $k = 1, \ldots, i$, by the main result in [1], $\phi_{A=Y}(\mathbf{r}) = \mathbf{r}_1(R_1) \otimes \ldots \otimes \mathbf{r}_i(R_i) \otimes \phi_{A=Y}(\mathbf{r}(R_j \cup \ldots \cup R_n))$.

Theorem 1 indicates our method is *sound*.

For example, nesting the set ADE of variables as the single variable G using the joint distribution $\mathbf{r}(R)$ in Figure 1 is the nested relation

$$\phi_{G=\{A,D,E\}}(\mathbf{r}),$$

depicted in Figure 2.

On the other hand, in practice the joint distribution $\mathbf{r}(R)$ is usually represented as the Markov network

$$\mathbf{r}_1(ABD) \otimes \mathbf{r}_2(ABC) \otimes \mathbf{r}_3(ACE) \otimes \mathbf{r}_4(BCF).$$

Our General Nest algorithm first applies the selective reduction algorithm to remove those tables that need not be joined. In this case, $R_4 = BCF$ is deleted. The Markov join

$$\mathbf{r}_1(ABD) \otimes \mathbf{r}_2(ABC) \otimes \mathbf{r}_3(ACE)$$

of the remaining tables is computed, as illustrated in Figure 4. As the set ADE of variables has been gathered into a single table, the nest operator is now applied giving

$$\phi_{G=\{A,D,E\}}(\mathbf{r}_1(ABD) \otimes \mathbf{r}_2(ABC) \otimes \mathbf{r}_3(ACE)),$$

as shown in Figure 5. It can be verified that

$$\mathbf{r}_4(BCF) \otimes \phi_{G=\{A,D,E\}}(\mathbf{r}_1(ABD) \otimes \mathbf{r}_2(ABC) \otimes \mathbf{r}_3(ACE))$$

is the same nested relation as $\phi_{G=\{A,D,E\}}(\mathbf{r})$, namely,

$$\phi_{G=\{A,D,E\}}(\mathbf{r}) = \mathbf{r}_4(BCF) \otimes \phi_{G=\{A,D,E\}}(\mathbf{r}_1(ABD) \otimes \mathbf{r}_2(ABC) \otimes \mathbf{r}_3(ACE)).$$

B	C	G				$p(BCG)$
		A	D	E	$p(ADE)$	
0	0	0	0	0	0.1	0.5
		0	0	1	0.1	
		1	2	2	0.8	
		A	D	E	$p(ADE)$	
0	1	0	0	1	1.0	0.2
		A	D	E	$p(DE)$	
1	0	0	1	0	0.5	0.3
		0	1	1	0.5	

Figure 5. The nested relation $\phi_{G=\{A,D,E\}}(\mathbf{r}_1(ABD) \otimes \mathbf{r}_2(ABC) \otimes \mathbf{r}_3(ACE))$, where $\mathbf{r}_1(ABD) \otimes \mathbf{r}_2(ABC) \otimes \mathbf{r}_3(ACE)$ is the relation in Figure 4.

5 Conclusion

Xiang [9] explicitly states that more work needs to be done on our *granular* probabilistic networks [6]. Here we have extended [1, 2] by coarsening an *arbitrary* set of variables without any restrictions. Theorem 1 establishes the *correctness* of our approach.

References

[1] Butz, C.J. and Wong, S.K.M. A local nest property in granular probabilistic networks, *Proceedings of the Fifth Joint Conference on Information Sciences*, **1**, 158–161, 2000.

[2] Butz, C.J., Hamilton, H.J. and Yao, H. A non-local coarsening result in granular probabilistic networks, *Proceedings of the Ninth International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing* (to appear), 2003.

[3] Koller, D. and Pfeffer, A. Object oriented Bayesian networks, *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, 302–313, 1997.

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

[5] Tarjan, R.E. and Yannakakis, M. Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs, *SIAM Journal on Computing*, 13(3):566–579, 1984.

[6] Wong, S.K.M. and Butz, C.J. Contextual weak independence in Bayesian networks, *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, 670–679, 1999.

[7] 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*, A, 30(6): 785–805, 2000.

[8] Wong, S.K.M. and Butz, C.J. Constructing the dependency structure of a multi-agent probabilistic network, *IEEE Transactions on Knowledge and Data Engineering*, 13(3):395–415, 2001.

[9] Xiang, Y. *Probabilistic Reasoning in Multi-Agent Systems: A Graphical Models Approach*, Cambridge Publishers, 2002.