XIV Congresso Ibero Latino-Americano de Métodos Computacionais em Engenharia

Instituto de Pesquisas Tecnológicas
São Paulo - SP - Brasil
1 a 3 de dezembro de 1993

UMA ESTRATÉGIA PARA REDUÇÃO DO CUSTO DA APLICAÇÃO DA PROPAGAÇÃO LOCAL DE INFORMAÇÃO EM SISTEMAS BASEADOS EM VALUAÇÕES

Sandra Sandri
Instituto Nacional de Pesquisas Espaciais (INPE)
Caixa Postal 515
CEP 12201-970 - São José dos Campos - SP
sandri@lac.inpe.br

SUMÁRIO

Apresentamos aqui uma estratégia capaz de reduzir o custo computacional da aplicação da propagação local da informação em sistemas baseados em conhecimento que usam valuações como modelo de representação da informação. Ela consiste em um pré-processamento que determina quais elementos da base são irrelevantes em relação a uma dada consulta. Esta estratégia aproxima sistemas baseados em valuações, onde é possível garantir a correção dos resultados, aos sistemas baseados em regras, onde somente uma parte da base de conhecimentos é manuseada.

1. Introdução

Nas últimas duas décadas temos assistido o desenvolvimento de um grande número de sistemas baseados em conhecimento, dentre os quais destacam-se os sistemas baseados em regras. Nestes sistemas combina-se uma extrema facilidade na obtenção do conhecimento dos especialistas com um importante poder computacional. No entanto, os mecanismos para a propagação da informação geralmente utilizados nestes sistemas são baseados na hipotése falaciosa de que os elementos básicos de conhecimento fornecidos por um especialista são independentes entre si. O uso de tais mecanismos pode levar a resultados que não são necessariamente corretos quando os dados são imperfeitos, i.e. quando representados dentro de algum modelo de incerteza.

Uma alternativa é o uso do modelo de propagação local da informação em sistemas baseados em valuações (Shafer e Shenoy, 1988). Com este mecanismo garante-se a correção dos resultados, porém apesar de não necessariamente incorrer-se numa explosão computacional, seu custo pode ser bastante vultuoso. Neste trabalho propomos um procedimento de pré-processamento que, ao identificar as informações irrelevantes em relação a uma dada consulta, nos permitirá descartá-las, reduzindo assim o custo do processamento, conservando no entanto da correção dos resultados.

O trabalho está dividido da seguinte maneira. Na Seção 2 apresentamos o formalismo básico usado neste trabalho e na Seção 3 apresentamos nossa estratégia. A Seção 4 traz a conclusão.

2. FORMALISMO DE BASE

Neste trabalho estamos interessados em sistemas baseados em conhecimento cujas bases podem ser codificadas na forma $\mathcal{K} = \{V_f \ / \ f \subset \mathcal{X}\}$, onde $\mathcal{X}$ é um conjunto de variáveis, f é um suconjunto de $\mathcal{X}$ e V_f é uma valuação (Shafer e Shenoy 1986) num dado modelo de incerteza. Valuações estendem o conceito do valor de uma variável para acomodar a imperfeição da informação. Na Teoria da Evidência (também conhecida como Teoria de Dempster-Shafer) (Shafer 1976) uma valuação é uma função de credibilidade $Cr : 2^{\Omega_h} \rightarrow [0,1]$, definida por $Cr(A) = \sum\{m(B) \ / \ B \subset A, B \neq \emptyset\}$, onde $m : 2^{\Omega_h} \rightarrow [0,1]$ é uma função de distribuição de massa que aloca a evidência que se refere estritamente a A e não a qualquer subconjunto de A. A função de credibilidade Cr tem como dual a função de plausibilidade Pl, definida por $Pl(A) = 1 - Cr(\neg A)$. Neste trabalho uma valuação V_h será representada por um *corpo de evidência* sobre h, que é formado pelo conjunto de pares $(A, m_h(A))$, tal que $A \in 2^{\Omega_h}$, e $m(A) > 0$ (m_h é a função de distribuição de massa sobre h).

Seja $\mathcal{X}$ um conjunto finito de variáveis x_i tomando valores em Ω_{x_i}, e $h \subset \mathcal{X}$ uma variável n-dimensional em $\Omega_h = \Omega_{x_{h1}} \times \ldots \times \Omega_{x_{hn}}$, $x_{hj} \in h$. Se $\mathcal{K}$ é uma base de conhecimento cujos elementos básicos de conhecimento são modelados através do uso de variáveis $h \subset \mathcal{X}$, então $\mathcal{K}$ é estruturado por um hipergrafo $\mathcal{H} = (S, \mathcal{X})$, $S \subseteq 2^{\mathcal{X}}$, onde cada hiperarco $h \in S$ é formado por um grupo de variáveis, e é tal que dois hiperarcos estão relacionados se eles partilham um grupo de variáveis. A Figura 2.1a traz um hipergrafo para $\mathcal{X} = \{x_i \ / \ 1 \leq i \leq 10\}$ e $S = \{x_3, x_6, x_8, x_1x_2, x_3x_7, x_3x_8, x_6x_9, x_7x_8, x_8x_{10}, x_1x_3x_4, x_1x_3x_7\}$.

Nos sistemas baseados em valuações, a inferência dos valores requeridos por uma consulta é feita através de operações de marginalização, extensão e combinação de valuações realizados ao longo de uma estrutura contruída a partir de seu hipergrafo. O operador de *marginalização* projeta a valuação V_g sobre uma variável $g \subset \mathcal{X}$, para uma valuação $V_g^{\downarrow f}$ sobre o domínio de referência de f, $f \subset g$. $V_g^{\downarrow f}$ é chamada de *marginal* de V_g sobre f. O operador de *extensão* projeta a valuação V_g sobre $g \subset \mathcal{X}$, para a valuação $V_g^{\uparrow h}$ sobre o domínio de referência de h, $h \supset g$. $V_g^{\uparrow h}$ é chamada de extensão de V_g sobre h. Por exemplo, seja $V_{x1x2} = \{(\neg\{1\}+\{a\}, .9), (\Omega_{x1} \times \Omega_{x2}, .1)\}$, onde $\Omega_{x1} = \{a,b\}$ e $\Omega_{x2} = \{1,2\}$, uma valuação descrita pela Teoria da Evidência. A marginalização de V_{x1x2} para a variável x_1 é $V_{x1x2}^{\downarrow x1} = \{(\Omega_{x1}, 1)\}$ e a extensão desta última valuação para x_1x_2 é $(V_{x1x2}^{\downarrow x1})^{\uparrow x1x2} = \{(\Omega_{x1} \times \Omega_{x2}, 1)\}$. É importante notar que neste caso particular $(V_{x1x2}^{\downarrow x1})^{\uparrow x1x2} \neq V_{x1x2}$.

O operador de *combinação* agrega duas valuações definidas sobre o mesmo universo ; se V_g e V_h são valuações sobre g e h respectivamente, então $V_g \otimes V_h = V_g^{\uparrow g \cup h} \otimes V_h^{\uparrow g \cup h}$ é uma valuação sobre $g \cup h$ representando a combinação de V_g e V_h. O operador de combinação na Teoria da Evidência é a Regra de Combinação de Dempster (Shafer 1976) : $m3(A) = 1/k \sum\{m_1(B)m_2(C) / B \quad C = A\}$, $k = 1 - m_3(\varnothing)$, onde m_i é a função de distribuição de massa de produzida pela fonte i.

O valor correto de uma variável k dadas as informações de uma base de conhecimento $\mathscr{K}$ sobre $\mathscr{H}$, usando um modelo de incerteza qualquer, é a marginalização para o universo de k da valuação global $V_S = \otimes\{V_f{}'^S / f \in S\}$. No entanto, o custo de obtenção da valuação global é proibitiva. Shafer e Shenoy (1986) (ver também (Shafer e Shenoy 1988)) provaram que se o modelo de incerteza adotado em $\mathscr{K}$ obedece os axiomas A1, A2 e A3 dados abaixo então a inferência feita através do mecanismo de *propagação local da informação* produz resultados corretos dentro daquele modelo de incerteza.

Axiom A1 (*Comutatividade e associatividade da combinação*) :
Suponhamos que V_g, V_h, V_k são valuações sobre g, h, e k respectivamente.
Então $V_g \otimes V_h = V_h \otimes V_g$ e $V_g \otimes (V_h \otimes V_k) = (V_g \otimes V_h) \otimes V_k$.
Axiom A2 (*Consonância da marginalização*) :
Suponhamos que $f \subset g \subset h$ e que Vf, Vg, V_h são valuações sobre f, g e h respec.
Então $V_h^{\downarrow f} = (V_h^{\downarrow g})^{\downarrow f}$
Axiom A3 (*Distributividade da marginalização sobre a combinação*) :
Suponha que V_g é uma valuação sobre g e que V_h é uma valuação sobre h.
Então $(V_g \otimes V_h)^{\downarrow g} = V_g \otimes (V_h{}^{\downarrow g \quad h})$

O processo de propagação local da informação consiste em identificar grupos de valuações que devem necessariamente ser combinadas conjuntamente, e propagar a informação localmente entre esses grupos. No primeiro passo deriva-se uma estrutura conhecida como árvore de Markov (tambem referenciadas como "Joint Trees" em literatura mais recente) a partir do hipergrafo subjacente a base.

Uma *árvore de Markov* $T = (N, E)$, $N \subset 2^{\mathscr{X}}$, $E \subset N \times N$, é um grafo conexo acíclico, tal que se f e g são dois vértices distintos de N, e $x \in \mathscr{X}$ é uma variável que pertence a ambos f e g, então x também pertence a todo vértice no caminho entre f e g. Em outras palavras, se uma variável x pertence a quaisquer dois vértices g e h, então existe um e somente um caminho p ligando h e g em T, e x pertence a todos os vértices ao longo de p (ver Fig.2.1b).

Uma árvore de Markov $T = (N,E)$ sobre $\mathscr{X}$ é dita ser representativa de um hipergrafo $\mathscr{H} = (S, \mathscr{X})$ se para cada hiperarco $s \in S$, existe um vértice $n \in N$ tal que $s \subset n$. Esta árvore preserva as inter-relações de $\mathscr{H}$, pois se dois hiperarcos estão relacionados em $\mathscr{H}$, então existem vértices $h' \supset h$ e $g' \supset g$ em N, tais que ou $h' = g'$, ou h' e g' são conectados por um caminho (único) em E. As árvores de Markov representativas de um dado hipergrafo podem ser derivadas através da aplicação de operadores especiais sobre S, que ou adiciona vértices de $2^{\mathscr{X}}$, ou substitui um vértice $s \in S$, por um vértice $t \in 2^{\mathscr{X}}$, tais que $s \subset t$. Informações específicas sobre processos de derivação de tais árvores, também chamados de *cobertura de hipergrafos*, podem ser encontrados em (Lianwen 1988), (Kong 1986), (Mellouli 1988). A árvore de Markov T da Figura 2.1b é representativa do hipergrafo apresentado na Figura 2.1a. Na criação de T os hiperarcos x_3x_7, x_3x_8 e x_7x_8 foram aglutinados (garantindo a consistência com a definição), gerando o nó $f = x_3x_7x_8$ com $V_f = V_{x_3x_7} \otimes V_{x_3x_8} \otimes V_{x_7x_8}$.

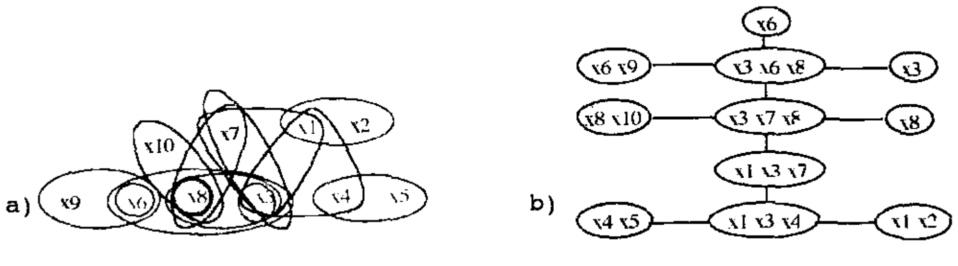

Fig.2.1a - Hipergrafo $\mathcal{H}$ em $\mathcal{X} = \{x_1,\ldots,x_9\}$.
2.1b - Árvore de Markov representativa de $\mathcal{H}$.

Suponhamos que deseja-se derivar o valor da variável k a partir de $\mathcal{K} = \{V_f \ / \ f \subset \mathcal{X}\}$. Seja T = (N,E) uma árvore de Markov representativa de $\mathcal{H}$, e seja h um nó de T, tal que $k \subset h$. O processo de *propagação local da informação* consiste em tomar h como nó raiz e propagar a informação a partir dos nós folha até h. Seja M o conjunto de nós folha. Um ciclo do processo de propagação consiste em três passos:

1) *Marginalização:* Toma-se cada nó n_i de M e marginaliza-se V_{ni} na direção de seu nó pai n_w gerando as valuações $V_{ni}^{\downarrow ni \cap nw}$.

2) *Extensão:* Cada valuação $V_{ni}^{\downarrow ni \cap nw}$ é estendida para $(V_{ni}^{\downarrow ni \cap nw})^{\uparrow nw}$.

3) *Combinação:* Para cada n_w, V_{n_w} e todas as $(V_{ni}^{\downarrow ni \cap nw})^{\uparrow nw}$ referentes a seus filhos são combinadas. Os nós n_w passam então a compor M.

O processo de inferência consiste pois em usar a operação de combinação no menor dos universos possíveis, e propagar o resultado de um nó para outro através da intersecção de seus universos. Elimina-se desta forma tanto perdas de informação devidas à marginalização quanto problemas decorrentes da eventual falta de idempotência do operador de combinação, e ao mesmo tempo reduz-se o custo da computação da valuação global. Os axiomas acima são satisfeitos dentro da Teoria da Evidência (Shafer e Shenoy 1988) e em vários outros modelos matemáticos (ver (Sandri 1993) para uma revisão).

Como uma árvore de Markov T = (N,E) é uma estrutura de grafo não orientada, cada arco $(f,g) \in E$ pode ser usado para propagar informação de f para g, e de g para f (obviamente, após o início do processo de propagação local somente um sentido é usado em cada arco). Embora o termo propagação local da informação refira-se a árvores de Markov em geral, o processo de propagação é na verdade efetuado sobre árvores de Markov orientadas, já que neste processo utiliza-se $T_h = (N,E_h)$, onde N é um conjunto de nós, E_h um conjunto de arcos orientados sobre N tomando-se h como nó raiz. Uma árvore de Markov $T = (N,E)$ define uma família de $|N|$ árvores de Markov orientadas, uma para cada nó de N tomado como raiz.

Embora o modelo de propagação local apresentado acima seja o mais geral, outros modelos foram implementados para casos específicos (ver (Sandri 1993) para uma revisão).

3. Estratégia para poda de árvores de Markov orientadas

Shenoy e Shafer (1990) mostram que a derivação das marginais para todos os nós na árvore de Markov é menos que três vezes o custo da derivação da marginal para um único nó. Para tanto basta transformarmos cada nó num processador que armazena, processa e transmite informação. Isto nos permite calcular as marginais para todas as variáveis em $\mathcal{K}$. No entanto, uma das vantagens de sistemas baseados em conhecimento é a flexibilidade na aquisição e modificação da informação. Neste caso é provavelmente inútil armazenar-se as marginais em todos os nós, já que eles deverão ser recalculados de tempos em tempos. Além disso, é importante observar que na maior parte das vezes estamos interessados em deduzir o valor de somente um pequeno número de variáveis. Assim, pode ser mais razoável desistir do custo da derivação do valor para todas as variáveis na base, se podemos obter em troca uma redução no custo da derivação do valor de nossas variáveis objetivo, através da eliminação na base de toda informação que seja irrelevante em relação a estas variáveis.

Parte da informação numa árvore de Markov Orientada pode ser eliminada, caso o modelo de incerteza subjacente a $\mathcal{K}$ obedeça o seguinte axioma (ver (Sandri, 1991) e (Cano et al, 1991)):

Axioma B *(Existência da identidade)* :
Seja V_f uma valuação sobre f. Então existe uma valuação V_{fe} sobre f tal que $\forall\, V_f$, $V_f \otimes V_{fe} = V_{fe} \otimes V_f = V_f$. V_{fe} é chamada de valuação identidade de f, ou a valuação vácua de f.

V_{fe} é completamente não informativa, ou em outras palavras, a partir de V_{fe} tudo o que podemos saber de f é que seu valor está certamente contido em Ω_f. O axioma B é por exemplo respeitado na Teoria da Evidência : para qualquer variável f em $\mathcal{X}$, sua valuação vácua é $V_{fe} = \{(\Omega_f, 1)\}$.

A marginalização (ou projeção) de uma valuação para um referencial menor será dita ser vácua se ela produz uma valuação vácua. Seja V_h uma valuação sobre h e $V_h^{\downarrow f}$ a marginalização de V_h para f, $f \subset h$. Definimos $\beta(V_h^{\downarrow f})$ como a proposição "a marginalização da valuação V_h para f é vácua", que será verdadeira se $V_h^{\downarrow f} = V_{fe}$. O conjunto $B(V_h) \subset 2^h$ reúne todas as variáveis contidas em h, para as quais a marginalização de V_h produz uma valuação vácua, i.e. $f \in B(V_h)$ sse $\beta(V_h^{\downarrow f})$ é verdadeira. Assumimos aqui que $\varnothing \in B(V_f)$, $\forall$ f $\in \mathcal{X}$.

Seja $\mathcal{F}$ com $\mathcal{H} = (S, \mathcal{X})$, $S = \{x_3, x_6, x_8, x_1x_2, x_3x_7, x_3x_8, x_6x_9, x_7x_8, x_8x_{10}, x_1x_3x_4, x_1x_3x_7\}$, e com os conjuntos $B(V_f)$ dados na Tabela 3.1. Suponhamos que $V_{x_3x_7x_8}$, obtido com a fusão de $V_{x_3x_7}$, $V_{x_3x_8}$ e $V_{x_7x_8}$, seja tal que obtenhamos $B(V_{x_3x_7x_8}) = \{\varnothing, \{x_3\}, \{x_7\}, \{x_3, x_7\}\}$. A Figura 3.1a mostra uma árvore para $\mathcal{K}$, com as marginalizações vácuas indicadas ao lado das flexas.

Tabela 3.1

Exemplo de conjuntos de marginalizações vácuas

$B(V_{x3x8})$ =	$\{\emptyset, \{x3\}, \{x8\}\}$
$B(V_{x3x7})$ =	$\{\emptyset, \{x3\}, \{x7\}\}$
$B(V_{x7x8})$ =	$\{\emptyset, \{x7\}, \{x8\}\}$
$B(V_{x1x3x4})$ =	$(\emptyset, (x1\}, \{x3\}, \{x4\}, \{x1, x3\}, \{x1, x4\}, \{x3, x4\}\}$
$B(V_{x4x5})$ =	$\{\emptyset, \{x4\}, \{x5\}\}$
$B(V_{x1x2})$ =	$\{\emptyset, \{x1\}, \{x2\}\}$
$B(V_{x1x3x7})$ =	$\{\emptyset, \{x1\}, \{x3\}, \{x7\}, \{x1, x3\}, \{x1, x7\}, \{x3, x7\}\}$
$B(V_{x6x9})$ =	$\{\emptyset, \{x6\}, \{x9\}\}$
$B(V_{x8x10})$ =	$\{\emptyset, \{x8\}, \{x10\}\}$
$B(V_{x6x8})$ =	$\{\emptyset, \{x3\}, \{x6\}, \{x8\}, \{x3, x6\}, \{x3, x8\}, \{x6, x8\}\}$
$B(V_{x3})$ =	$\{\emptyset\}$
$B(V_{x6})$ =	$\{\emptyset\}$
$B(V_{x8})$ =	$\{\emptyset, \{x8\}\}$

Seja $T_n = (N, E_n)$, uma árvore de Markov orientada obtida a partir de uma árvore de Markov T com a designação de n como nó raiz. Quando queremos propagar a informação num nó folha f para um nó vizinho g, devemos projetar a valuação em f para $f \cap g$, e então combiná-la com a informação em g. No entanto, se a projeção é vácua, este passo de propagação como um todo pode ser considerado como inútil, já que o resultado final será a informação já contida em g. Logo, se o modelo de incerteza subjacente a $\mathcal{K}$ obedece os axiomas A_1, A_2, e A_3 e também o axioma B, podemos podar um nó folha f de T_n, se $f \cap g \in B(V_f)$, onde g é o nó vizinho de f. O uso sucessivo deste mecanismo nos nós folha de T_n nos permitirá podá-la, e assim reduzir o custo efetivo da propagação. Por exemplo, dada a árvore da Figura 3.1a tendo a variável x8 como objetivo, a eliminação de valuações vácuas resultaria na árvore de Markov orientada da Figura 3.1b. Neste caso a informação seria propagada por apenas 4 arcos ao invés dos 10 arcos do procedimento padrão.

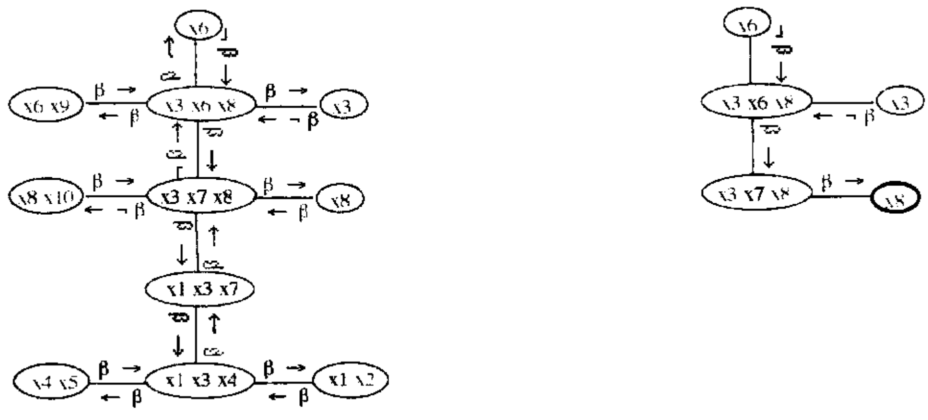

Fig.3.1a - Arvore de Markov

3.1b - Arvore de Markov orientada com raiz x_8 podada.

Uma decorrência dos axiomas A1, A2 e A3 é que, para uma base de conhecimento qualquer $\mathcal{K}$ com hipergrafo $\mathcal{H}$, o resultado da aplicação da propagação local da informação é o mesmo sobre não importa qual árvore de Markov derivada a partir de $\mathcal{H}$. Portanto se construirmos uma árvore de Markov de tal forma que possamos aumentar o número de nós com marginalizações vácuas na direção da raiz, obteremos uma redução do custo da propagação local. Isso é especialmente interessante se considerarmos que em aplicações reais uma grande parte do conhecimento na base é irrelevante em relação a uma dada consulta. Este é por exemplo o caso de informações derivadas a partir de regras (ver Sandri 1993)).

A estratégia que propomos aqui constrói inicialmente uma árvore de Markov somente para as informações gerais na base, e acrescenta as informações específicas de forma a maximizar a poda de valuações vácuas. Em consequência, se as informações gerais da base não foram modificadas, como é frequentemente o caso, outras consultas poderão ser feitas sem o custo de obtenção da árvore inicial.

Nossa estratégia consiste basicamente em :

a) Construir uma árvore de Markov somente para os nós na base de conhecimento que concernem informações de caráter geral,

b) Determinar uma árvore de Markov orientada T_n relativa à variável objetivo n,

c) Inserir os nós contendo informação específica em T_n, o mais próximo possível da raiz,

d) Podar a árvore resultante, e propagar finalmente a informação.

O processo é dividido em 6 fases.

Fase 1 : Particionamento da base e construção da árvore de Markov

Dividimos $\mathcal{K}$ em duas partes : a *base de conhecimento não informada* $\mathcal{NK} = \{V_f \in \mathcal{K} \ / \ |f| > 1, \ B(V^f) = 2^f - f\}$, e a *base de conhecimento informada* $\mathcal{IK} = \mathcal{K} - \mathcal{NK} - \{V_f \in \mathcal{K} \ / \ f \in B(V^f)\}$. O conjunto $\mathcal{K} - \mathcal{NK} - \mathcal{IK}$ é desprezado pois contém somente as valuações vácuas. Para podermos garantir que a estrutura que será derivada por nossa estratégia seja consistente com a definição de árvores de Markov devemos impor três restrições sobre os elementos de conhecimento na base :

1) Para cada valuação V_g em $\mathcal{IK}$, deve existir uma valuação V_h em $\mathcal{NK}$ tal que $g \subset h$.

2) Sejam h e g duas variáveis quaisquer em $\mathcal{IK}$. Então $h \cap g = \emptyset$. Caso contrário, substituimos V_h e V_g em $\mathcal{IK}$ por $V_{h \cup g} = V_h \otimes V_g$.

3) Sejam h e g duas variáveis quaisquer em $\mathcal{NK}$. Então $h \cap g \neq h$. Caso contrário, substituimos V_h e V_g em $\mathcal{NK}$ por $V_{h \cup g} = V_h \otimes V_g$.

Fase 2 : Determinação dos arcos vácuos

Nesta fase tomamos cada arco $e = (f,g)$ em E e rotulamos cada um de seus sentidos com ou $\beta(V_f^{\downarrow f \cap g})$ ou $\neg\beta(V_g^{\downarrow f \cap g})$; i.e. verificamos as marginalizações vácuas em T.

Fase 3 : Particionamento da árvore de Markov

Seja $l(h, g)$ o comprimento do caminho entre h e g em T, e $l(h, G) = \inf_{g \in G} l(h, g)$ o comprimento do menor caminho entre f e um dos nós em G. Seja $R = \{r \ / r \in N, \ x \in r\}$ um conjunto de nós em T contendo a variável objetivo x. Nesta fase particionamos primeiramente o conjunto de nós N - R pela menor distância entre um nó em N - R e os nós em R. Formalmente, a partição $\mathcal{P}_c$ é composta das classes $c_i = \{h \in N - R \ / \ l(h, R) = i\}$, $1 \leq i \leq |N - R|$. De maneira similar, o conjunto R é particionado pela menor distância entre um nó em R e os nós em N - R. A partição $\mathcal{P}_d$ é composta das classes $d_i = \{f \in R \ / \ l(h, N - R) = i\}$, $1 \leq i \leq |R|$. A classe d_1 é chamada de classe raiz. As partições $\mathcal{P}_c$ e $\mathcal{P}_d$ são utilizadas para dar uma orientação em todos os nós de T exceto aqueles em d_1. A Figura 3.2a traz a árvore de Markov derivada pra $\mathcal{NF}$, particionada usando x_8 como variável objetivo.

Fase 4 : Determinação de nós indispensáveis

Quando desejamos propagar a informação em f para seu nó pai g, devemos antes de mais nada marginalizar a valuação em f para $f \cap g$. Se o arco (f, g) tem o rótulo $\neg\beta(V_f{}^{\downarrow f \cap g})$, então a informação em f não pode ser desprezada, porque ela é possivelmente informativa. Nesta fase, marcamos como indispensáveis todos da árvore que são ou informados, ou localizados entre um nó informado e um nó no nível d_1. Iniciamos o processo nas folhas (tanto nos níveis c_i quanto d_i), e progredimos na árvore até só restar o nível d_1 a ser examinado. Marcamos então como indispensáveis todos os nós em d_1 que tem vizinhos indispensáveis nos outros níveis. Finalmente todos os nós em d_1 localizados entre dois nós indispensáveis de d_1 são também marcados como indispensáveis. Denotamos por N_I o conjunto de todos os nós indispensáveis de T ; *todos* estes nós devem ser levados em consideração no processo de propagação, independentemente da informação armazenada em $\mathcal{SF}$. Na Figura 3.2a somente o nó $\{x_3, x_7, x_8\}$ é marcado como indispensável.

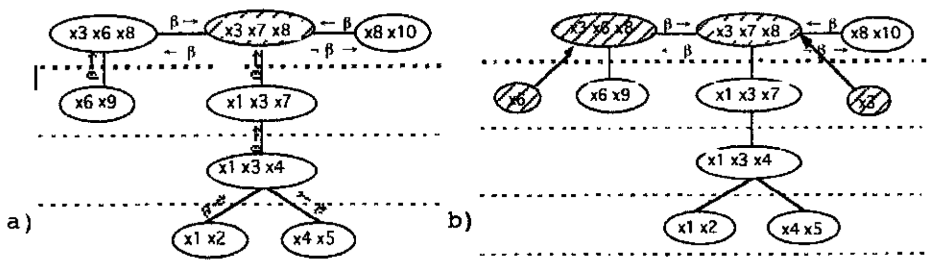

Fig.3.2a - Após as fases 3 e 4
3.2b - Após as fases 5 e 6

Fase 5 : Inserção dos nós informados

Nesta fase inserimos cada hiperarco f contido na base $\mathcal{SF}$ na árvore de Markov T = (N, E) da seguinte maneira. Seja $Y(f) = \{g \ / \ g \in N, \ g \supset f\}$ o conjunto de nós em T que contem f. Seja $i = \inf_{g \in Y(f)} l(g, d_1)$, a menor distância entre Y(f) e a classe d_1. Primeiramente, tentaremos ligar f a um nó indispensável em T. No entanto, se isto não for possível, ligamos f a qualquer nó $Y(f) \cap c_i$, ou $Y(f) \cap d_i$, o mais próximo possível do nível d_1. Seja k o nó escolhido. Então marcamos f e k como indispensáveis, bem como todos os nós no caminho entre k e um nó em d_1. A Figura 3.2b ilustra a inserção dos nós de $\mathcal{IF}'$ na árvore derivada para $\mathcal{NK}$.

Fase 6 : Propagação da informação

A árvore de Markov será finalmente caracterizada ao criarmos um nó y formado unicamente pela variável objetivo e ligá-lo a um nó raiz qualquer de $d_1 \cap N_I$. A informação será então propagada através dos nós em N_I, a partir dos nós folhas até g. No nosso exemplo, a árvore resultante será aquela mostrada na Figura 3.1b.

4. Conclusão

Apresentamos aqui uma estratégia para reduzir o custo do processo de propagação local da informação. Ela se aplica a qualquer teoria que satisfaça os axiomas de Shafer e Shenoy (Shafer e Shenoy 1988), e o axioma da existência da identidade (Sandri 1991) (Cano et al 1991).O último axioma nos permite verificar se uma informação na base é irrelevante e podendo assim ser ignorada no processo de propagação.

Nossa estratégia consiste em construir uma árvore de Markov para os nós não informados em $\mathcal{K}$, e transformá-la gradativamente numa árvore orientada com a inserção dos nós informados. Como na maior parte do tempo os nós informados numa base de conhecimento concernem conhecimento específico à consulta, a separação da base numa parte informada e noutra não informada é equivalente àquela existente em sistemas baseados em regras, onde informações de caráter geral estão contidos na base de regras, e as restantes na base de fatos. Nestes sistemas somente uma parte da base é explorada : aquela correspondente à consulta. Aqui, também não exploramos necessariamente (no sentido de combinar) todas as informações na base, mas temos a certeza de que toda a informação desprezada não interferiria no resultado final. Além disso, uma boa parte do trabalho realizado em uma consulta pode ser usado com outras informações específicas e com outras variáveis objetivo. Na verdade, se a informação geral na base não for modificada, não há necessidade de determinar-se uma nova árvore de Markov ou de recalcular-se os conjuntos de marginalizações vácuas.

Referências

- Cano J.E., Delgado M. & Moral S. (1991) *Propagation of Uncertainty in Dependence Graphs. Symbolic et Quantitative Approaches to Uncertainty*, Proc. European Conf. ECSQAU, Marseille, 1991.

- Kong A. (1986), *Multivariate Belief Functions and Graphical Models*, Doctoral Dissertation, Department of Statistics, Harvard University.

- Lianwen Z. (1988), *Studies on Finding Hypertree Covers for Hypergraphs*, Working Paper N° 198, School of Business, The University of Kansas, Lawrence.

- Mellouli K. (1988) *On the Propagation of Belief in Networks Using the Dempster-Shafer Theory of Evidence*, Working Paper N°.196, School of Business, The University of Kansas, Lawrence.

Shafer G. (1976), *A Mathematical Theory of Evidence*, Princeton University Press.

- Sandri S.A. (1991), *La combinaison de l'information incertaine et ses aspects algorithmiques*, Doctoral Thesis, Université Paul Sabatier, Toulouse, France.

- Sandri, S. A. (1993), "Local Propagation of Information on Directed Markov Trees", in *Intelligent Systems with Uncertainty*, (B. Bouchon-Meunier, L. Valverde, R Yager, eds), Elsevier, to appear in 1993.
- Shafer G. e Shenoy P.P. (1986), "Propagating Belief Functions with Local Computation", *IEEE Expert*, 1(3), 43-51.
- Shafer G. e Shenoy P.P. (1988), *Local Computation in Hypertrees* Working Paper N° 201, School of Business, The University of Kansas, Lawrence.
- Shenoy P.P. e Shafer G. (1990), "Axioms for Probability and Belief-function Propagation", In *Uncertainty in Artificial Intelligence 4*, (Schachter R.D., Levitt T.S., Kanal L.N., Lemmer J.F., eds), Elsevier Science Publishers B.V. (North Holland).