

DERIVADAS EM REDES BAYESIANAS USANDO ELIMINAÇÃO DE VARIÁVEIS

MATHIAS JUAN PERAZZO*, FÁBIO GAGLIARDI COZMAN†

**Rua Itapaiuna 1165 casa 68, 05707-001
São Paulo, SP, Brasil*

†*Escola Politécnica, Universidade de São Paulo
Av. Prof. Mello Moraes, 2231, 05508-900
São Paulo, SP, Brasil*

Emails: mathiasp@engineer.com, fgcozman@usp.br

Abstract— Bayesian networks are extensively used in artificial intelligence, pattern recognition and system identification. An important operation is the differentiation of Bayesian networks; that is, the computation of derivatives for the parameters of a Bayesian network. This paper presents a new method for such a computation: we present an algorithm based on the variable elimination method, and discuss the advantages and applications of this new algorithm.

Keywords— Bayesian network, derivation, variable elimination, sensitivity analysis, expert systems

Resumo— Redes Bayesianas são extensivamente usadas em inteligência artificial, reconhecimento de padrões e identificação de sistemas. Uma operação importante é a diferenciação de redes Bayesianas; isto é, o cálculo de derivadas de parâmetros de uma rede Bayesianas. Este artigo apresenta um novo método para este cálculo: apresentamos um algoritmo baseado no método de eliminação de variáveis, e discutimos as vantagens e aplicações deste novo algoritmo.

Palavras-chave— Redes Bayesianas, derivação, eliminação de variáveis, análise de sensibilidade, sistemas especialistas

1 Introdução

Redes Bayesianas são modelos voltados à tomada de decisão e à manipulação de incertezas, com aplicações em inteligência artificial, reconhecimento de padrões, identificação de sistemas (Jensen, 1996). São modelos gerais e compactos, com capacidade para representar situações estáticas e dinâmicas. Sistemas especialistas com redes Bayesianas encontram uso em diversas áreas: diagnóstico, prognóstico e planejamento médico do tratamento (Coupé et al., 1999), tendo como exemplo o sistema especialista com rede Bayesianas Alarm para diagnóstico médico (Beinlich et al., 1989); interpolação automática de dados brutos de sondas interplanetárias (Stutz et al., 1998); sistemas especialistas como o Pathfinder e o Munnin (Jensen, 1996), etc.

Este artigo foca na *diferenciação de parâmetros de redes Bayesianas*. Derivadas de parâmetros de redes Bayesianas são de grande utilidade em pelo menos duas aplicações: realização de análise de sensibilidade em redes Bayesianas, e otimização de parâmetros por métodos de gradiente. Este trabalho tem por objetivo apresentar um método para se obter derivadas em redes Bayesianas baseado no *algoritmo de eliminação de variáveis*. Esse método é de simples implementação computacional e de fácil aprendizado, cujos resultados em nada ficam devendo a algoritmos existentes que envolvem conceitos gráficos.

A Seção 2 contém um breve resumo da teoria de redes Bayesianas e do algoritmo de elimi-

nação de variáveis aplicado em redes Bayesianas. A Seção 3 por sua vez faz uma análise dos métodos existentes para cálculo de derivadas em redes Bayesianas, e a Seção 4 apresenta o método proposto neste trabalho. A Seção 5 contém um exemplo resolvido passo a passo pelo método proposto. A Seção 6 indica duas possíveis aplicações para a teoria aqui apresentada, e a Seção 7 conclui o artigo e indica trabalhos futuros.

2 Redes Bayesianas e Eliminação de Variáveis

Assume-se neste artigo que todas as variáveis aleatórias têm um número finito de possíveis valores. Um conjunto de variáveis é escrito em negrito (por exemplo, \mathbf{X}). A notação $\mathbf{X} \setminus \mathbf{Y}$ indica o conjunto de variáveis que pertence a \mathbf{X} mas não a \mathbf{Y} . A notação $\sum_{\mathbf{X}} f(\mathbf{X}, \mathbf{Y})$ indica que todas as variáveis em \mathbf{X} são eliminadas (somando) da função $f(\mathbf{X}, \mathbf{Y})$. A densidade de probabilidade de X é $p(X)$: $p(x)$ é o valor da probabilidade do evento $\{X = x\}$. A densidade de probabilidade de X condicional a valores de Y é $p(X|Y)$.

Uma rede Bayesianas representa uma densidade de probabilidade sobre um conjunto de variáveis \mathbf{X} (Pearl, 1998; Jensen, 1996). A densidade é especificada através de um grafo direcional acíclico. Cada nó nesse grafo representa uma variável aleatória X_i em \mathbf{X} . O conjunto de variáveis $pa(X_i)$ denota os pais de X_i , $ch(X_i)$ denota os filhos de X_i . E os pais dos filhos de X_i que não sejam eles próprios filhos são as “esposas”

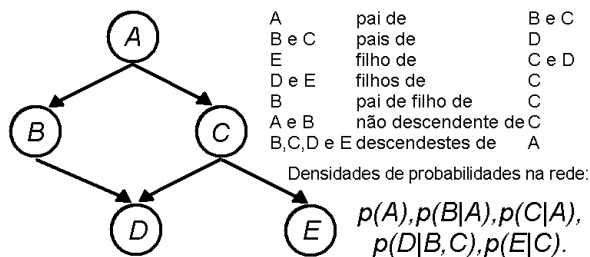


Figura 1: Exemplo de rede Bayesiana.

$spo(X_i)$. A Figura 1 ilustra relações em uma rede. Uma rede Bayesiana atende à condição de Markov: cada nó é condicionalmente independente de seus não descendentes não pais dados seus pais. Essa condição leva a uma distribuição conjunta de probabilidade única (Pearl, 1998):

$$p(\mathbf{X}) = \prod_i p(X_i | pa(X_i)).$$

Dado uma rede Bayesiana, o evento E denota as *evidências* ou *observações* na rede. Por exemplo $E = \{X_1 = x_{12}, X_3 = x_{31}\}$ fixa os valores das variáveis X_1 e X_3 .

Inferência com redes Bayesianas envolve normalmente o cálculo da probabilidade marginal a *posteriori* para um conjunto de variáveis questionadas \mathbf{X}_q (Pearl, 1998). A probabilidade a *posteriori* de \mathbf{X}_q dado E é:

$$P(\mathbf{X}_q | E) = \frac{P(\mathbf{X}_q, E)}{P(E)} = \frac{\sum_{\mathbf{X} \setminus \{\mathbf{X}_q, \mathbf{X}_E\}} P(\mathbf{X})}{\sum_{\mathbf{X} \setminus \mathbf{X}_E} P(\mathbf{X})}.$$

Eliminação de variáveis é um algoritmo para inferência exata em redes Bayesianas (Dechter, 1996; Zhang and Poole, 1996). Na sua essência o método simplesmente elimina (somando) repetitivamente as variáveis. Dada uma rede Bayesiana de variáveis \mathbf{X} , evidência E e variáveis questionadas \mathbf{X}_q , inferir $p(\mathbf{X}_q | E)$ envolve normalmente apenas um subconjunto das densidades associadas com a rede. Se a densidade $p(X_i | pa(X_i))$ for necessária para produzir uma inferência então X_i é uma *variável requisitada*; existem algoritmos polinomiais para encontrar todas as variáveis requisitadas para uma inferência. O conjunto de variáveis requisitadas é expresso por \mathbf{X}_R .

Seja N o número de variáveis requisitadas que não são observadas e não estão em \mathbf{X}_q . (Cozman, 2000b) demonstra que:

$$p(\mathbf{X}_q, E) = \sum_{X_N} \cdots \sum_{X_2} \left(\prod_{\substack{X_i \in \\ X_R \setminus \{X_1, ch(X_1)\}}} p(X_i | pa(X_i)) \right) \times \left(\sum_{X_1} \prod_{\substack{X_j \in \\ \{X_1, ch(X_1)\}}} p(X_j | pa(X_j)) \right).$$

Sendo que se pode definir a seguinte densidade não normalizada:

$$p(ch(X_1) | pa(X_1), spo(X_1)) = \sum_{X_1} \left(\prod_{X_j \in ch(X_1)} p(X_j | pa(X_j)) \right).$$

Ou seja, pode-se eliminar a variável X_1 do problema. Repetindo-se sucessivamente para X_2 até X_N no final teremos algumas (pelo menos uma) densidades para \mathbf{X}_q . Através da multiplicação dessas densidades e normalizando o resultado obtemos $p(\mathbf{X}_q | E)$.

A ordenação das variáveis a serem eliminadas é arbitrária, mas diferentes ordenações levam a diferentes cargas computacionais. Existem vários métodos heurísticos conhecidos que produzem ordenações eficientes (Zhang and Poole, 1996).

Uma vez que a ordenação das variáveis esteja definida, o algoritmo de eliminação de variáveis cria estruturas de dados chamadas *buckets*. Cada *bucket* contém uma variável a ser eliminada X_i e todas as densidades que contenham a variável do *bucket*. A seqüência de *buckets* criada é chamada de *árvore de buckets*.

Suponha agora que se deseje computar a densidade de probabilidade marginal para cada variável numa rede Bayesiana. (Cozman, 2000b) generaliza o algoritmo de eliminação de variáveis; através do armazenamento de resultados intermediários no processo de eliminação, essa generalização consegue produzir inferências para todas as variáveis na rede de forma conjunta.

3 Diferenciação de Redes Bayesianas

Existem vários métodos para derivar redes Bayesianas, ou seja, para computar o valor da derivada de uma probabilidade com respeito a algum parâmetro da rede Bayesiana. A seguir serão citados os métodos mais importantes.

3.1 Força bruta

Este primeiro método não calcula exatamente derivadas, mas produz indiretamente o seu valor. Para cada probabilidade de interesse é investigado um número de variações em relação ao valor inicial. Obtém-se dessa forma a variação em relação

à condição inicial. Esse método tem a desvantagem de consumir muito esforço computacional, pois cada variação em valores deve ser acompanhada de uma inferência.

3.2 Fracional linear

O método proposto por (Coupé et al., 1999) é baseado na observação de que a relação entre uma probabilidade de interesse e um parâmetro de uma Rede Bayesiana pode ser expressa como um quociente de duas funções lineares.

Esse método requer algumas definições preliminares:

- 1) Define-se x_{ik} como um valor de uma variável X_i e π uma combinação arbitrária de valores do conjunto de pais $pa(X_i)$;
- 2) Definindo também $y = p(a|E)$, onde a é o valor da variável de interesse A na derivada ($A \in \mathbf{X}$).
- 3) Assume-se que numa derivação, em se variando o parâmetro $x = p(x_{ik}|\pi)$ todas as outras probabilidades $p(x_{il}|\pi)$ são co-variadas de acordo, pelo escalonamento da razão entre as massas de probabilidade restantes. Mais formalmente, seja o domínio da variável X_i $dom(X_i) = \{x_{i1}, \dots, x_{im}\}$, $m \geq 1$. Os parâmetros $p(x_{il}|\pi)$, $k \neq l$, são funções de x :

$$p(x_{il}|\pi)(x) = \begin{cases} x & \text{se } l = k \\ p(x_{il}|\pi)(x) \frac{1-x}{1-p(x_{ik}|\pi)} & \text{se } l \neq k \end{cases}$$

com $x = p(x_{ik}|\pi) < 1$.

Temos que (Coupé et al., 1999; U.Kjærulff and van der Gaag, 2000):

Teorema 1 *Seja p a função de probabilidade definida por uma rede Bayesiana sobre um conjunto de variáveis \mathbf{X} . Seja $y = p(a|E)$ e $x = p(x_{ik}|\pi)$ como indicados anteriormente. Então,*

$$y = \frac{p(a, E)(x)}{p(E)(x)} = \frac{\alpha x + \beta}{\gamma x + \delta},$$

onde α , β , γ e δ são constantes com relação a x .

As constantes dessa função determinam diretamente as derivadas de interesse. Determinar essas constantes requer apenas três inferências na rede, que podem ser realizadas com qualquer algoritmo de inferência.

3.3 Método por árvore de junção

(U.Kjærulff and van der Gaag, 2000) apresentam um método para derivação de redes Bayesianas em sua representação *árvore de junção* (Jensen, 1996; Pearl, 1998). O método é baseado na idéia de que, numa *árvore de junção*, a expressão para $p(a, E)$ e $p(E)$ em função de x pode ser obtida do potencial de um clique contendo tanto a variável x como seus pais. O seguinte teorema detalha como os coeficientes podem ser calculados.

Teorema 2 *Seja p a função de probabilidade definida por uma rede Bayesiana e seja T a representação da rede na sua forma de árvore de junção. Seja $y = p(a|E)$ e $x = p(x_{ik}|\pi)$ como indicados anteriormente. Suponha que, em T , uma propagação partindo da extremidade tenha sido realizada em direção ao clique que contém a variável de interesse A ; suponha que subsequente uma propagação em direção à extremidade tenha sido realizada desse clique com o valor a para A . Agora, seja Q um clique em T contendo tanto a variável X_i e seus pais $pa(X_i)$; seja $\phi_Q = p(Q, a, E)$ o potencial do clique Q depois das propagações citadas acima. Então, $p(a|E)(x) = \alpha x + \beta$ onde:*

$$\alpha = \frac{\sum_{Q:x_{ik}, \pi} \phi_Q}{p(x_{ik}|\pi)} - \sum_{l \neq k} \frac{\sum_{Q:x_{il}, \pi} \phi_Q}{1 - p(x_{ik}|\pi)} \quad (1)$$

$$\beta = \sum_{l \neq k} \frac{\sum_{Q:x_{il}, \pi} \phi_Q}{1 - p(x_{ik}|\pi)} + \sum_{Q:pa(X_i) \neq \pi} \phi_Q \quad (2)$$

Suponha agora que a evidência E tenha sido processada em T por uma propagação do topo em direção a raiz e subsequente propagação da raiz em direção ao topo. Seja $\phi_Q^* = p(Q, E)$ o potencial do clique Q depois das propagações citadas acima. Então, $p(E)(x) = \gamma x + \delta$ onde:

$$\gamma = \frac{\sum_{Q:x_{ik}, \pi} \phi_Q^*}{p(x_{ik}|\pi)} - \sum_{l \neq k} \frac{\sum_{Q:x_{il}, \pi} \phi_Q^*}{1 - p(x_{ik}|\pi)} \quad (3)$$

$$\delta = \sum_{l \neq k} \frac{\sum_{Q:x_{il}, \pi} \phi_Q^*}{1 - p(x_{ik}|\pi)} + \sum_{Q:pa(X_i) \neq \pi} \phi_Q^* \quad (4)$$

O teorema acima fornece a base para o método de computar os coeficientes das funções que expressam a probabilidade de interesse $y = p(a|E)$ em termos de todos os possíveis parâmetros x .

Este último método requer apenas três propagações na rede (na sua representação *árvore de junção*) para estabelecer todas as derivadas para uma probabilidade posterior marginal específica. Note que o método *Fracional Linear* precisa de três inferências na rede para cada parâmetro.

4 Um Novo Método Baseado em Eliminação de Variáveis

A proposta deste trabalho é a de se adaptar o método de derivação proposto por (U.Kjærulff and van der Gaag, 2000), ao algoritmo de inferência em redes Bayesianas proposto por (Cozman, 2000b) que generaliza eliminação de variáveis.

O método de (U.Kjærulff and van der Gaag, 2000) é todo baseado e construído em cima de redes Bayesianas na sua representação na forma de *árvore de junção*. Ou seja, utiliza complexos conceitos e métodos da teoria de grafos. Porém, o

algoritmo de generalização de eliminação de variáveis também pode ser interpretado utilizando-se ferramentas de teoria de grafos. O resultado básico aqui é que toda árvore de *buckets* é também uma árvore de cliques (os cliques da árvore de triangulação são induzidos pelo ordenamento das variáveis) e portanto pode ser visto como uma *árvore de junção*.

Tendo isso em mente, propõe-se o seguinte algoritmo que adapta a derivação do método por *árvore de junção* na estrutura de *buckets* de *eliminação de variáveis*. O método é bastante similar a uma inferência em árvore de *buckets*, mas contém alguns passos adicionais para cálculo de valores usados em derivadas:

1. Gere um ordenamento para as N variáveis que são necessárias para a inferência, que não observadas e não questionadas. (A probabilidade de interesse na derivação A deve ser a variável questionada \mathbf{X}_q na eliminação de variáveis)
2. Coloque todas as densidades da rede em um conjunto de densidades.
3. Para i de 1 a N :
 - a. Crie uma estrutura de dados B_i , chamada *bucket*, contendo:
 - A variável X_i , chamada variável do *bucket*;
 - Todas as densidades que contenham a variável do *bucket*, chamado de densidade do *bucket*;
 - b. Multiplique as densidades em B_i . Armazene a densidade resultante não normalizada em B_i ; a densidade é chamada de *cluster* de B_i .
 - c. Elimine (somando) X_i do cluster B_i . Armazene a densidade resultante não normalizada em B_i ; a densidade é chamada de separador de B_i .
 - d. Coloque o separador do *bucket* em um conjunto contendo densidades.
4. No final do processo colete as densidades que contém a variável questionadas num bucket B_q . Multiplique as densidades em B_q (e normalize o resultado para obter $p(A|E)$ com as condições iniciais da rede).
A propagação agora para atualizar os *buckets* parte da raiz, *bucket* B_q que por definição já está atualizado após o passo 4, em direção ao topo.
5. Para i de N a 1:
 - a. Normalize o *cluster* de B_i em relação à variável do *bucket*, obtendo-se a densidade normalizada $p(X_i|\mathbf{S}_i, E_i)$;
 - b. Peça ao filho de B_i que forneça $p(\mathbf{S}_i|E_i)$.

Esta densidade de probabilidade substitui o separador de B_i ;

c. Multiplique $p(\mathbf{S}_i|E_i)$ e o *cluster* normalizado do *bucket* $p(X_i|\mathbf{S}_i, E_i)$, obtendo $p(X_i, \mathbf{S}_i|E_i)$. Esta densidade de probabilidade substitui o cluster de B_i .

6. Compute os coeficientes γ e δ , usando as equações (3) e (4), para todos os parâmetros relevantes, localmente por clique/*bucket*.
7. Realize outra propagação saindo da raiz em direção ao topo de clique/*bucket* B_q , com a evidência adicional $A = a$.
8. Compute os coeficientes α e β , usando as equações (1) e (2), para todos os parâmetros relevantes, localmente por clique/*bucket*.

5 Exemplo

Nesta seção a teoria proposta é aplicada num exemplo simples. Para tanto é utilizada a rede da figura 1 com os seguintes valores de probabilidades (considerando que todas as variáveis são booleanas):

$p(a) =$	0,20	$p(d c, b) =$	0,80
$p(c a) =$	0,20	$p(d \neg c, b) =$	0,80
$p(c \neg a) =$	0,05	$p(d c, \neg b) =$	0,80
		$p(d \neg c, \neg b) =$	0,05
$p(b a) =$	0,80	$p(e c) =$	0,80
$p(b \neg a) =$	0,20	$p(e \neg c) =$	0,60

Também é assumido como evidência que a variável E é verdadeira. A probabilidade de interesse é $p(c|e)$. Vamos estudar a derivada em relação a três probabilidades: $p(e|\neg c)$, $p(c|\neg a)$ e $p(a)$.

Aplicando os conceitos de d-separação a rede é reduzida a três variáveis (A , C e E), gerando uma *árvore de buckets* com dois *buckets*. A variável do *bucket* B_1 é A e a variável questionada C está no *bucket* B_q .

Utilizando o algoritmo de eliminação de variáveis inicia-se pelo *bucket* B_1 . A variável do *bucket* é A e as densidades do *bucket* são $p(A)$ e $p(C|A)$. O *cluster* de B_1 é obtido através da multiplicação das densidades e elimina-se (somando) a variável A obtendo o separador S_1 .

$$\begin{aligned} \phi_1(A, C) \Rightarrow & \begin{aligned} a, c &= 0,04 \\ a, \neg c &= 0,16 \\ \neg a, c &= 0,04 \\ \neg a, \neg c &= 0,76 \end{aligned} \\ S_1 = \sum_A \phi_1 & \Rightarrow \begin{aligned} c &= 0,08 \\ \neg c &= 0,92 \end{aligned} \end{aligned}$$

Passa-se agora para o *bucket* B_q . A variável do *bucket* é a variável questionada C e as densidades do *bucket* são $p(C)$ e o separador S_1 . Multiplicando as densidades obtém-se a variável questionada

nada C .

$$\begin{aligned} c &= 0,064 \\ \neg c &= 0,552 \end{aligned}, \text{ normalizando obtemos}$$

$$\begin{aligned} c &= 0,103896 \\ \neg c &= 0,896104 \end{aligned}$$

Generalizando eliminação de variáveis é realizada a segunda propagação, desta vez partindo da raiz em direção ao topo e com o valor obtido para a variável questionada C .

Primeiro normaliza-se o *cluster* de B_1 em relação a variável A e depois multiplica-se o resultado por $p(C|E)$ obtido na primeira propagação que substitui o valor do separador S_1 .

$$\begin{aligned} \phi_1(A, C) \Rightarrow \quad & a, c = 0,5 \\ & a, \neg c = 0,173913 \\ & \neg a, c = 0,5 \\ & \neg a, \neg c = 0,826087 \end{aligned}$$

multiplicando por $p(C|E)$

$$\begin{aligned} a, c &= 0,0519481 \\ a, \neg c &= 0,155844 \\ \neg a, c &= 0,0519481 \\ \neg a, \neg c &= 0,740260 \end{aligned}$$

Pode-se obter o valor de $p(A|E)$ eliminando a variável C

$$\sum_C \phi_1^* \Rightarrow \begin{aligned} a &= 0,207792 \\ \neg a &= 0,792208 \end{aligned}$$

Calcula-se os coeficientes γ e δ para as três probabilidades em estudo:

$$\begin{aligned} p(e|\neg c) \Rightarrow \quad & \gamma = 0,92 \\ & \delta = 0,064 \\ p(c|\neg a) \Rightarrow \quad & \gamma = 0,259740 \\ & \delta = 0,987013 \\ p(a) \Rightarrow \quad & \gamma = 0,0487013 \\ & \delta = 1,99026 \end{aligned}$$

Realizando uma última propagação de B_q para B_1 com a evidência adicional $C = c$.

$$\begin{aligned} \phi_q \Rightarrow \quad & c = 0,064 \\ & \neg c = 0 \\ \phi_1 \Rightarrow \quad & a, c = 0,0519481 \\ & a, \neg c = 0 \\ & \neg a, c = 0,0519481 \\ & \neg a, \neg c = 0 \end{aligned}$$

Finalmente, computa-se os coeficientes α e β para as três probabilidades em estudo:

$$\begin{aligned} p(e|\neg c) \Rightarrow \quad & \alpha = 0 \\ & \beta = 0,064 \\ p(c|\neg a) \Rightarrow \quad & \alpha = 1,03896 \\ & \beta = 0,0519481 \\ p(a) \Rightarrow \quad & \alpha = 0,194805 \\ & \beta = 0,168831 \end{aligned}$$

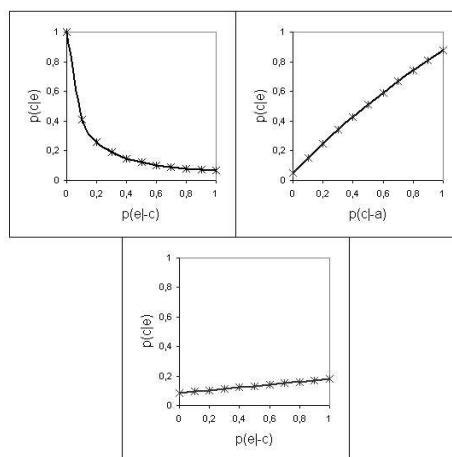


Figura 2: Curvas das derivadas.

Para comprovar a eficiência do método aqui utilizado foi realizada uma comparação com uma derivação através de força bruta (para tanto foi utilizado o programa JavaBayes). As respostas coincidiram. Para visualizar os resultados e a comparação realizada foram geradas as três curvas (utilizando-se os coeficientes obtidos) e em cada curva foram marcados os dez pontos da análise por força bruta. A Figura 2 traz esses resultados.

6 Aplicações

A seguir serão citadas duas aplicações diretas do método aqui proposto.

6.1 Análise de sensibilidade de redes Bayesianas

Redes Bayesianas são usualmente construídas com o auxílio de especialistas no assunto (*domain experts*). Construir a parte quantitativa, ou seja, atribuir valores para as probabilidades condicionais de todas as variáveis representadas na rede, é considerada a tarefa mais difícil (Coupé et al., 1999).

Dada uma rede Bayesiana e uma inferência de interesse, nem todo valor de probabilidade na rede vai requerer o mesmo nível de precisão para que a inferência tenha uma precisão satisfatória. Algumas probabilidades têm um impacto maior na rede do que outras. Dessa forma, quando se deseja refinar os valores de probabilidades (determinar o nível de precisão requerido para as diversas probabilidades condicionais da rede), é recomendado realizar uma análise de sensibilidade (Coupé et al., 1999; Chan and Darwich, 2001). Através de uma análise de sensibilidade descobrem-se quais são as variáveis que exercem maior influência (valores de probabilidades para os quais o sistema mostrou ser mais sensível). Devido ao tempo limitado e custoso de especialistas, o refinamento

(ou otimização) pode ser focado nessas probabilidades.

Quando se deseja descobrir qual é a influência que a variação de uma probabilidade em estudo (ou de um grupo de probabilidades) tem em uma probabilidade de interesse, se deseja na realidade saber a derivada da rede. Uma derivada onde a saída (ou valor da derivada) é o valor da probabilidade de interesse e se está derivando em relação à probabilidade em estudo.

6.2 Otimização Local de Parâmetros

Em muitas ocasiões o valor exato de uma probabilidade $p(X_i|pa(X_i))$ pode não ser conhecido. Isso ocorre quando o especialista construindo a rede tem informações vagas ou insuficientes; ou quando um conjunto de especialistas diverge sobre os parâmetros de uma rede; ou quando dados experimentais são coletados e não são suficientes para estimar um único valor de probabilidades. Nesses casos, uma solução possível é considerar intervalos de probabilidades para os valores $p(X_i|pa(X_i))$, e realizar inferências a partir desses intervalos (Cano et al., 1993; Cozman, 2000a; Fagiouli and Zaffalon, 1998). A realização de inferências requer então a otimização de valores de probabilidades com as restrições prescritas por intervalos. Em tais casos, um método de inferência importante é a otimização local por método do gradiente, a qual exige o cálculo de derivadas em cada uma de suas iterações. Os resultados apresentados nesse artigo podem então ser utilizados.

7 Conclusões

O algoritmo apresentado neste trabalho para se obter derivadas em redes Bayesianas é de fácil entendimento e implementação. Diferente de métodos atuais, como o (U.Kjærulff and van der Gaag, 2000), este algoritmo não utiliza conceitos de grafos, tais como triangulação e cliques, evitando dessa forma a utilização de complexos métodos da teoria de grafos. Ao invés disso, este algoritmo se foca apenas em densidade de probabilidades.

Aplicações deste algoritmo são de grande utilidade. Métodos eficientes para análise de sensibilidade tem uma função importante tanto nas fases de aquisição de conhecimento como na fase de validação na construção manual de modelos de redes Bayesianas. A otimização local de valores de probabilidade tem aplicações quando parâmetros não são conhecidos com precisão absoluta, por serem baseados em conhecimento limitados ou em dados experimentais finitos.

O passo seguinte será estender os resultados aqui obtidos para uma derivação de n parâmetros. Ou seja, deseja-se descobrir o efeito não mais de uma probabilidade em estudo, mas sim estudar o

efeito de um conjunto de probabilidades (duas ou mais) na probabilidade de interesse.

Referências

- Beinlich, I., Suermondt, H. J., Chavez, R. M. and Cooper, G. F. (1989). The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks, *Second European Conference on Artificial Intelligence in Medicine*, p247-256.
- Cano, J., Delgado, M. and Moral, S. (1993). *An axiomatic framework for propagating uncertainty in directed acyclic networks*. International Journal of Approximate Reasoning.
- Chan, H. and Darwich, A. (2001). *When do numbers really matter?* Conference on Uncertainty in Artificial Intelligence, p. 65-74.
- Coupé, V. M. H., van der Gaag, L. C. and Habbema, J. D. F. (1999). *Sensitivity Analysis of Decision-Theoretic Networks*, Holanda.
- Cozman, F. G. (2000a). *Credal networks*. Artificial Intelligence, 199-233.
- Cozman, F. G. (2000b). *Generalizing Variable Elimination in Bayesian Networks*, Atibaia, Brazil. Workshop on Probabilistic Reasoning in Artificial Intelligence.
- Dechter, R. (1996). *Bucket propagation: A unifying framework for probabilistic inference*. Conference on Uncertainty in Artificial Intelligence, p.211-219.
- Fagiouli, E. and Zaffalon, M. (1998). *2U: An exact interval propagation algorithm for polytrees with binary variables*. Artificial Intelligence 106(1)77:107.
- Jensen, F. V. (1996). *An Introduction to Bayesian Networks*, Springer Verlag.
- Pearl, J. (1998). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann.
- Stutz, J., Taylor, W. and Cheeseman, P. (1998). *AutoClass C - General Information*. NASA, Ames Research Center.
- U.Kjærulff and van der Gaag, L. C. (2000). *Making Sensitivity Analysis Computationally Efficient*, Holanda. Conference on Uncertainty in Artificial Intelligence.
- Zhang, N. L. and Poole, D. (1996). *Exploiting causal independence in Bayesian network inference*. Journal of Artificial Intelligence Research, p. 301-328.