

Uma Propriedade Simplificadora no Problema da Propagação de Evidências em Redes Bayesianas¹

Nelson Lopes Duarte Filho²

Tarcísio Pequeno³

Departamento de Informática, PUC/RJ. Rua Marquês de São Vicente, 225
22453 - Rio de Janeiro, RJ

Resumo

A fim de tornar tratável o problema da propagação de evidências em Redes Bayesianas, NP-difícil no caso geral, heurísticas que visam a simplificação da rede tem sido sugeridas. Em [7], por exemplo, é apresentado um método que consiste em identificar variáveis que são irrelevantes com respeito a propagação e eliminá-las da rede. No mesmo espírito, introduzimos aqui uma nova classe dessas variáveis, que denominamos parcialmente relevantes, e provamos um teorema que estabelece a possibilidade da sua eliminação. Uma heurística baseada no teorema é então apresentada.

Abstract

The propagation of evidences in a Bayesian Network is a NP-hard problem. In order to handle it, several heuristics introducing simplifications in the network have been suggested. In [7], for instance, a method is presented which consists in the identification of irrelevant variables and their elimination from the network. A new class of those variables, which we call partially relevant, is defined here. A theorem establishing the possibility of their elimination is, then, proved and a heuristic based on the theorem is presented.

¹ Trabalho parcialmente financiado pela IBM do Brasil e FINEP.

² Doutorando em Informática;
Em licença da Fundação Universidade do Rio Grande, RS;
Áreas de interesse: Inteligência Artificial, Raciocínio Automatizado, Raciocínio Impreciso e Aprendizado Automático.

³ Doutor em Teoria da Programação;
Em licença da Universidade Federal do Ceará, CE;
Áreas de interesse: Lógica e Computação, Inteligência Artificial, Raciocínio Automatizado e Raciocínio Impreciso.

1.Introdução

Redes Bayesianas são grafos dirigidos e acíclicos cujos vértices representam variáveis aleatórias e as arestas suas interdependências, quantificadas por probabilidades condicionais. Cada vértice contém as distribuições de probabilidades das variáveis nele representadas, condicionadas aos valores das variáveis que imediatamente o antecedem no grafo. Prestam-se sobremaneira para representar distribuições de probabilidades conjuntas, de uma forma compacta, explorando o conceito de independência condicional [1,2,3]. Exemplificando, a distribuição de probabilidades conjuntas de três variáveis A, B, C, conforme suas interdependências, podem ser descritas da maneira a seguir apresentada e representadas sob a forma das Redes Bayesianas mostradas na Figura 1.

a) Se A, B e C são totalmente dependentes entre si:

$$P(A,B,C) = P(A) \cdot P(B|A) \cdot P(C|A,B)$$

b) Se B e C são independentes dado A:

$$P(A,B,C) = P(A) \cdot P(B|A) \cdot P(C|A)$$

c) Se A, B e C são independentes:

$$P(A,B,C) = P(A) \cdot P(B) \cdot P(C)$$

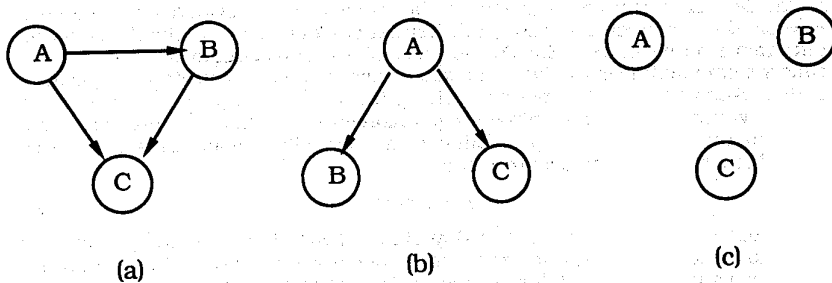


Figura 1

Existem procedimentos que aplicados a uma Rede Bayesiana são capazes de computar distribuições de probabilidades posteriores [4,5,6,7,8,9]. Ou seja, resolvem, por exemplo, problemas do tipo: "Qual a probabilidade de $A = a$ dado que $B = b$, nas Redes Bayesianas da Figura 1?". O processo desencadeado por esses procedimentos é chamado de propagação de evidências.

As Redes Bayesianas, juntamente com os procedimentos de propagação, mostram-se cômodos para a modelagem do raciocínio evidencial, e, particularmente, do diagnóstico, uma vez que o conceito de independência condicional é extremamente aderente aos conceitos de causa, influência, relevância e correlatos [10]. Por exemplo, após estruturar algum conhecimento médico sob a forma de uma Rede Bayesiana - os vértices representando patologias, sinais, sintomas, etc. -, estabelecer o diagnóstico de um paciente corresponde a propagar na rede as evidências nele

observadas e inspecionar as distribuições posteriores das patologias representadas.

Uma das inconveniências desse método de modelagem diz respeito ao fato do problema de propagação ser NP-difícil. Apenas para redes simplesmente conectadas (com apenas um caminho entre dois nós quaisquer) existem algoritmos polinomiais. Para minimizar esse problema, algumas heurísticas tem sido propostas. Entre as quais, as que procuram despovoar a rede, eliminando as variáveis irrelevantes com respeito a propagação [12]. Com essa filosofia, propõe-se neste artigo uma heurística, alicerçada em um teorema, que elimina as variáveis aqui chamadas *parcialmente relevantes*, visando diminuir a dificuldade do problema de propagação.

Na seção 2 são apresentadas definições e a terminologia a serem utilizadas, na seção 3 o Teorema da Relevância Parcial, na seção 4 a heurística de eliminação e, finalmente, na seção 5 as conclusões e extensões das idéias propostas.

2. Definições e Terminologia

A seguir, introduz-se algumas definições oriundas da Teoria dos Grafos [13], e a terminologia a ser adotada ao longo do artigo.

Um grafo dirigido acíclico (digrafo) $D(V,E)$ é um conjunto finito não vazio V (vértices) e um conjunto E (arestas) de pares ordenados de vértices distintos, sem ciclos. Uma sequência de vértices v_1, v_2, \dots, v_k , tal que $(v_j, v_{j+1}) \in E$, $1 \leq j < |V|$, é denominado caminho, $k - 1$ é o seu comprimento. Ciclo é um caminho v_1, \dots, v_k, v_{k+1} , sendo $v_1 = v_{k+1}$ e $k \geq 3$. Cada aresta (v, w) possui uma única direção, de v para w . Diz-se também que (v, w) é divergente a v e convergente a w . O grau de entrada de v é o número de arestas convergentes a v . Um vértice com grau de entrada zero é uma fonte. Todo vértice v do qual existe um caminho para w é denominado ancestral de w , se além disso for adjacente a w , denomina-se pai de w . Se forem retiradas as direções das arestas de um digrafo D obtém-se um grafo não direcionado (as arestas são pares não ordenados) G , chamado grafo subjacente a D . Com relação ao grafo subjacente podemos definir as seguintes propriedades: Um sub-grafo $G_2(V_2, E_2)$ de um grafo $G_1(V_1, E_1)$ é um grafo tal que $V_2 \subseteq V_1$ e $E_2 \subseteq E_1$. Se além disso, G_2 possuir toda aresta (v, w) de G_1 , tal que ambos v e w estão em V_2 , então G_2 é o sub-grafo induzido pelo sub-conjunto de vértices V_2 . Um grafo $G(V,E)$ é denominado conexo quando existe caminho entre cada par de vértices de G , caso contrário G é desconexo. Um corte de vértices é um subconjunto minimal de vértices $V' \subseteq V$ cuja remoção o desconecta. Denomina-se conectividade de vértices C_v a cardinalidade do menor corte de vértices de G . Sendo K um inteiro positivo, diz-se que um grafo é K -conexo em vértices quando a sua conectividade em vértices for $\geq K$. Denomina-se componentes bi-conexas do grafo aos sub-grafos maximais que sejam 2-conexas em vértices. Cada componente bi-conexo é chamado bloco do grafo. Um vértice v é denominado articulação quando sua remoção desconecta G . As articulações, portanto, são os pontos de corte que determinam os blocos do grafo. O grafo bloco-articulação de um grafo G é definido como o grafo B cujos vértices são os blocos de G e, sendo b_i, b_j vértices de B , (b_i, b_j) será

aresta de B caso b_i e b_j compartilhem uma articulação em G (adjacentes). Os grafos bloco-articulação são árvores, qualquer que seja G. Chama-se fator de convergência de um bloco b_i a uma articulação a_j ao número de arestas de b_i convergentes a a_j .

Redes Bayesianas são grafos dirigidos e acíclicos cujos vértices armazenam as distribuições de probabilidades das variáveis aleatórias que eles representam, condicionadas aos valores dos seus pais (os vértices fonte armazenam distribuições "a priori").

A seguinte notação será adotada ao longo do artigo:

- $R_i(V_i, E_i)$: Rede Bayesiana com vértices V_i e arestas E_i . Cada vértice será chamado $X_1, X_2, \dots \subseteq V_i$, e seus conjuntos $S_1, S_2, \dots \subseteq V_i$;
- $B(R_i)$: conjunto dos blocos de R_i com elementos B_1, B_2, \dots ;
- $I(R_i)$: conjunto das articulações de R_i com elementos I_1, I_2, \dots ;
- $F(X_i, S_i)$: conjunto dos pais de X_i que pertencem a S_i ;
- $C(I_i)$: conjunto dos blocos que possuem arestas convergentes a I_i com elementos C_{i1}, C_{i2}, \dots ;

3. Teorema da Relevância Parcial

Lema: Dada a Rede Bayesiana $R_1(V_1, E_1)$, se $R_2(V_2, E_2)$ é uma sub-rede conexa de R_1 , induzida por $V_2 \subseteq V_1$, com $B(R_2) \subseteq B(R_1)$, então a distribuição de probabilidades de V_2 , $P\{V_2\}$, pode ser representada por R_2 .

Prova:

Expressando $P\{V_1\}$ sob a forma do produto

$$P\{V_1\} = \prod_{X_i} P\{X_i | F(X_i, V_1)\} \quad (1)$$

e como

$$P\{V_2\} = \sum_{X_i \notin V_2} \prod P\{X_i | F(X_i, V_1)\} \quad (2)$$

se agruparmos os termos de (1) da seguinte maneira

$$\begin{aligned}
P\{V_1\} = & \prod_{X_i \in V_2, F\{X_i, V_1\} \subseteq V_2} P\{X_i | F\{X_i, V_1\}\} \cdot \\
& \prod_{X_j \in V_2, F\{X_j, V_1\} \subseteq V_2} P\{X_j | F\{X_j, V_1\}\} \cdot \\
& \prod_{X_k \in V_2, X_k \in A(V_2)} P\{X_k | F\{X_k, V_1\}\} \cdot \\
& \prod_{X_l \in V_2, X_l \in A(V_2)} P\{X_l | F\{X_l, V_1\}\}
\end{aligned} \tag{3}$$

podemos ver que (2) reduz-se a

$$\begin{aligned}
P\{V_2\} = & \prod_{X_i \in V_2, F\{X_i, V_2\} \subseteq V_2} P\{X_i | F\{X_i, V_2\}\} \cdot \\
& \prod_{X_j \in V_2, F\{X_j, V_2\} \subseteq V_2} P\{X_j | F\{X_j, V_2\}\}
\end{aligned} \tag{4}$$

que é exatamente a distribuição de probabilidades de V_2 representada por R_2 .

#

Teorema: Dada a Rede Bayesiana $R_1(V_1, E_1)$, $P\{S_1 | S_2\}$ pode ser calculada a partir de $R_2(V_2, E_2)$, menor sub-rede conexa de R_1 , induzida por $V_2 \subseteq V_1$, com $B(R_2) \subseteq B(R_1)$, que satisfaça $S_1 \cup S_2 \subseteq V_2$.

Prova:

$P\{S_1 | S_2\}$ pode ser calculada a partir de $P\{V_2\}$:

$$P\{S_1 | S_2\} = \frac{\sum_{X_i \in S_1 \cup S_2} P\{V_2\}}{\sum_{X_i \in S_2} P\{V_2\}} \tag{5}$$

e $P\{V_2\}$ pode ser representada por R_2 , conforme o Lema acima.

#

4. Heurística para Eliminação das Variáveis Parcialmente Relevantes

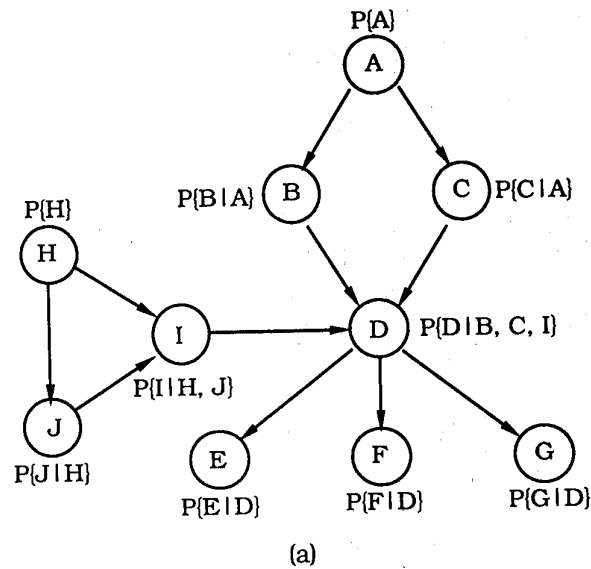
Dada uma Rede Bayesiana $R_1(V_1, E_1)$, sobre a qual se pretende propagar evidências, a seguinte heurística pode possibilitar uma economia no tempo de propagação:

Calcular e armazenar, apenas uma vez, quando a rede for construída, para cada articulação de R_1 , as distribuições de probabilidades:

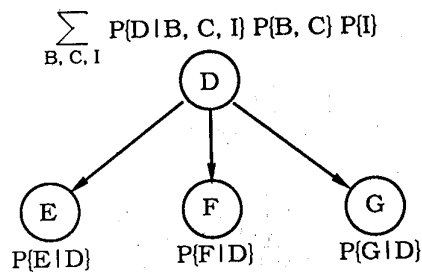
(6)

$P(F|I_i, C_{ij})$

Sempre que uma propagação para inferir $P(S_1|S_j)$ deva ser efetuada, determinar a menor sub-rede conexa $R_2(V_2, E_2)$ de R_1 , induzida por $V_2 \subseteq V_1$, com $B(R_2) \subseteq B(R_1)$, que satisfaça $S_1 \cup S_j \subseteq V_2$. E, para cada articulação pertencente a R_2 que possua arestas convergentes, oriundas dos blocos deixados de fora ($B(R_1) - B(R_2)$), calcular:



(a)



(b)

Figura 2

$$P\{I_1|F(I_1, V_2)\} = \sum_{C_{1k} \in BR_2} P\{I_1|F(I_1, V_1)\} \prod_{C_{1k} \in BR_2} P\{F(I_1, C_{1k})\} \quad (7)$$

e efetuar a propagação em R_2 .

Para se ter uma idéia do funcionamento da heurística, suponha a solução do problema de propagação "Qual a probabilidade de $D = d$ dado que $E = e$, $F = f$ e $G = g$, na rede da Figura 2(a)?" Após a aplicação da heurística, fica reduzido ao problema de propagação na rede da Figura 2(b).

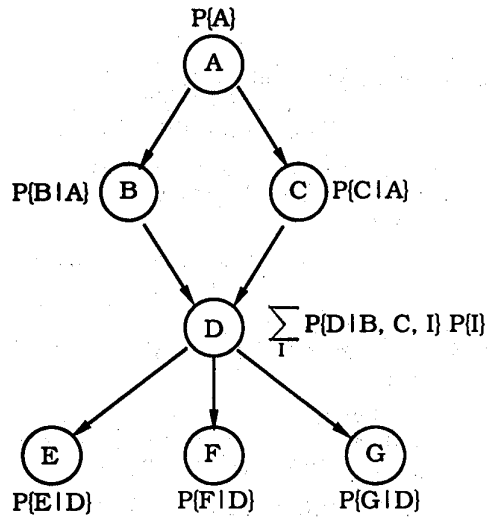


Figura 3

5. Conclusões

Foi apresentada uma heurística que visa diminuir a dificuldade do problema de propagação de evidências em Redes Bayesianas, que é fundamentalmente NP-difícil. Como pode ser visto no exemplo da seção 4, a heurística foi capaz de transformar um problema de propagação em uma rede multiconectada a um problema de propagação em uma rede simplesmente conectada, para o qual existem soluções por algoritmos polinomiais.

Cabe então, comentar sobre a complexidade dos algoritmos necessários a implementação da heurística. No que diz respeito a determinação dos componentes biconexos e articulações do grafo subjacente a rede, existem algoritmos $O(|V|+|E|)$ [13]. O problema maior surge no cálculo das probabilidades descritas por (6), que apesar de calculadas apenas uma vez, pode exigir um esforço inaceitável. Para solucionar esse problema, pode ser introduzido na heurística um regulador n , de tal forma que sejam calculadas apenas as distribuições de probabilidades $P\{F(I_i, C_{ij})\}$ para $|F(I_i, C_{ij})| \leq n$. Para calcular $P\{S_i|S_j\}$ numa rede $R_1(V_1, E_1)$ deve-se efetuar a

propagação na rede R_3 , obtida da menor sub-rede $R_2(V_2, E_2)$ de R_1 , induzida por $V_2 \subseteq V_1$, com $B(R_2) \subseteq B(R_1)$, que satisfaça $S_1 \cup S_j \subseteq V_2$, acrescida recursivamente dos blocos de R_1 que se conectam a R_2 com fator de convergência $\geq n$. Por exemplo, a aplicação da heurística com $n=1$, ao problema da Figura 2(a), o transforma no problema da Figura 3.

Bibliografia

- [1] Barlow, R. E., Pereira, C. A. de B., The bayesian operation and probabilistic influence diagrams (TR-ESRC 87-7), Department of Engineering and Operation Research, Berkeley, University of California, 1987;
- [2] Duarte Filho, N. L., Ferlin, C., Pequeno, T. H. C., Sistemas especialistas em diagnostico medico com capacidade para adquisicion de conocimiento, Dep. Inf., PUC/RJ, 1988 (Apresentado no Congresso Internacional INFORMATICA'88, Havana, Cuba);
- [3] Duarte Filho, N. L., Ferlin, C., Raciocinio evidencial e aprendizado automatico com o uso de Redes Bayesianas, Anais do VI Simp. Bras. em IA: 169-183, 1989;
- [4] Pearl, J., Fusion, propagation and structuring in belief networks, Artificial Intelligence 29: 241-288, 1986;
- [5] Shachter, R. D., Evidence absorption and propagation through evidence reversal, Proc. of the Fifth Workshop on Uncertainty in AI, Windsor, Ontario, 1989;
- [6] Lauritzen, S. L., Spiegelhalter, D. J., Local computations with probabilities on graphical structures and their application to expert systems, J. R. Statist. Soc. B 50, No. 2, 1988;
- [7] Pearl, J., Evidential reasoning using stochastic simulation of causal models, Artificial Intelligence 32: 245-257, 1987;
- [8] Duarte Filho, N. L., Ferlin, C., Protótipo para simulação de raciocínio evidencial com o uso de Redes Bayesianas, Exame de Qualificação, Dep. Inf., PUC/RJ, 1989;
- [9] Peot, M. A., Shachter, R. D., Fusion and propagation with multiple observations in belief networks, Dep. of Engineering-Economic Systems, Stanford University, 1989;
- [10] Pearl, J., Reverend Bayes in inference engines: A distributed hierarchical approach, Proc. Natl. Conf. on AI, 133-136, Pittsburgh, 1982;
- [11] Cooper, G. F., Probabilistic inference using belief networks is NP-hard, Report KSL-87-27, Medical Computer Science Group, Stanford University, 1987;
- [12] Shachter, R. D., An ordered examination of influence diagrams, Dep. of Engineering-Economic Systems, Stanford University, 1989 (to appear in Networks);
- [13] Szwarcfiter, J. L., Algoritmos e grafos: Uma introdução, Terceira Escola de Computação, Dep. Inf., PUC/RJ, Rio de Janeiro, 1982.