A/R+: Um Algoritmo para Inferência em Redes Credais.

José Carlos Ferreira da Rocha[1,2*], **Fábio Gagliardi Cozman**[1†]

[1]Laboratório de Telecomando e Tomada de Decisões
Departamento de Engenharia Mecatrônica
Escola Politécnica da Universidade de São Paulo
Av. Prof. Mello Moraes, 2231
05508-900 São Paulo, SP

[2]Departamento de Informática
Universidade Estadual de Ponta Grossa
Av. Gen. Carlos Cavalcanti, 4647
84013-900 Ponta Grossa, PR

`jrocha@uepg.br, fgcozman@usp.br`

Abstract. *This article presents the A/R+ algorithm for approximate inference in credal networks (a framework to represent imprecise probabilistic knowledge). Given a credal network, an inference aims at computing the upper and lower bounds for the marginal probability of each category of a random variable in the network. The proposed algorithm extends the A/R algorithm, an algorithm for approximate inference with interval-valued Bayes nets that can be applied to credal networks. Bounds obtained by A/R+ are significantly more precise than those produced by A/R.*

Resumo. *Este artigo apresenta o algoritmo A/R+ para inferência aproximada em redes credais (um esquema para representação de conhecimento probabilístico impreciso). O processo de inferência objetiva a computação dos limites superiores e inferiores da probabilidade marginal para cada categoria de uma variável aleatória da rede. O algoritmo proposto extende o algoritmo A/R, um algoritmo para inferência aproximada em redes Bayesianas com intervalos de probabilidade que pode ser aplicado sobre redes credais. As aproximações obtidas por A/R+ são significativamente mais precisas do que aquelas produzidas pelo algoritmo A/R.*

1. Introdução

O formalismo das *Redes Bayesianas* [J.Pearl, 1988] é um esquema de representação de conhecimento que tem sido empregado no desenvolvimento de aplicações que envolvem conhecimento incerto [D.Heckerman et al., 1995]. Uma rede Bayesiana $\mathbf{B}$ é definida pela dupla $(\mathbf{G}, \mathbf{P})$ onde: $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ é um grafo acíclico e direcionado no qual os nós em

*Bolsista CAPES/PICDT. Apoio HP Labs.
†Apoio CNPq e HP Labs, HP Brasil e Instituto de Pesquisa Eldorado.

$\mathbf{V} = \{V_1, ..., V_m\}$ simbolizam variáveis aleatórias e as arestas em $\mathbf{E} = \{E_1, ..., E_p\}$ indicam relações de dependência condicional direta; $\mathbf{P}$ é uma distribuição conjunta de probabilidades sobre as variáveis em $\mathbf{V}$. A topologia do grafo estabelece que toda variável é independente das demais variáveis da rede, exceto seus descendentes, se o estado de seus nós pais é conhecido[1]. Uma rede Bayesiana é uma polytree se $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ não apresenta ciclos mesmo quando a direção dos arcos é ignorada.

Uma rede Bayesiana é um esquema econômico para codificar distribuições conjuntas de probabilidades [E.Charniak, 1991]. Numa rede Bayesiana uma distribuição conjunta é codificada através de probabilidades condicionais. Isto é feito da seguinte maneira. Cada vértice V_i armazena uma tabela de probabilidades condicionais (TPC) com os valores de $p(V_i|pa(V_i))$, obtidas a partir de $\mathbf{P}$, sendo que $pa(V_i)$ denota os nós pais de V_i no grafo. Isto é, para cada instanciação conjunta k de $pa(V_i)$ existe na TPC uma entrada com a distribuição condicional $p(V_i|pa(V_i)_k)$. A distribuição conjunta original $\mathbf{P}$ pode ser recuperada através da expressão:

$$\mathbf{P}(\mathbf{V}) = \prod_{V \in \mathbf{V}, P(pa(V))>0} p(V|pa(V))\,. \tag{1}$$

Em geral, a execução de inferências sobre redes Bayesianas objetiva a determinação da probabilidade marginal para as categorias de uma variável $V_q \in \mathbf{V}$ dado um conjunto de evidências. Assim, se $\mathbf{O} \subseteq \mathbf{V}$ é o conjunto das variáveis observadas uma inferência objetiva computar $p(V_q|\mathbf{O})$ a partir de $\mathbf{P}$. Este processo é denominado *atualização de crença*.

Em algumas situações é útil considerar a presença de incerteza ou imprecisão com respeito ao valores nas TPCs. Este é o caso, por exemplo, quando: (a) deseja-se proceder uma análise de sensibilidade sobre uma rede Bayesiana[F.G.Cozman, 1997]; (b) o tempo e os recursos disponíveis não permitem a obtenção de um conjunto de parâmetros com a precisão desejada [J.P.Krause, 1998]; (c) os agentes envolvidos na engenharia de conhecimento discordam sobre os valores das TPCs [F.G.Cozman, 1999]. Nestes casos a escolha de um único conjunto de parâmetros para as TPCs pode ser uma estratégia inadequada.

O formalismo das *Redes Credais* propõe uma estratégia para abordar estas questões [F.G.Cozman, 2000a]. De forma análoga a uma rede Bayesiana, uma rede credal é um grafo acíclico e direcionado cujos vértices representam variáveis aleatórias. Também como em uma rede Bayesiana a topologia do grafo estabelece que, dados seus pais, um nó é independente dos demais nós da rede exceto seus descedentes. Contudo, numa rede credal uma variável pode ser associada não apenas a uma única TPC, mas a um ou mais conjuntos convexos de distribuições de probabilidade. Cada um destes conjuntos é denominado *conjunto credal* [I.Levi, 1980]. Assim, este formalismo usa conjuntos credais para representar incerteza e imprecisão com relação aos parâmetros numéricos de uma rede Bayesiana.

Dada uma variável aleatória em uma rede credal, um processo de inferência tem como objetivo computar os limites que restringem os valores de probabilidade que podem ser associados a cada uma das categorias da variável. Assim, para cada categoria

[1] O termo nó também será usado para indicar uma variável aleatória de uma rede Bayesiana/credal.

da variável selecionada estes limites, denominados *probabilidade superior* e *probabilidade inferior*, definem os extremos de um intervalo de probabilidades. A determinação dos valores exatos para estes limites é um problema NP-Completo mesmo em redes cuja topologia tem a forma de uma *polytree* [J.C.F.Rocha and F.G.Cozman, 2002].

Este artigo apresenta o algoritmo A/R+ para inferência aproximada em redes credais com topologia em *polytree*. É assumido que as relações de independência destacadas pela topologia das redes credais são de *independência forte* [I.Couso et al., 2000], um dos conceitos que podem ser utilizados para definir a relação de independência entre conjuntos credais. O algoritmo A/R+ implementa uma alteração no algoritmo A/R [B.Tessem, 1992]. O algoritmo A/R foi proposto por B.Tessem para inferência aproximada em redes Bayesianas cujas TPCs possuem intervalos de probabilidade, contudo também pode ser aplicado sobre redes credais. As aproximações obtidas por A/R+ são externas (contêm o intervalo exato), e substancialmente mais precisas do que aquelas calculadas por A/R quando este é aplicado sobre redes credais.

O artigo esta organizado da seguinte maneira. A seção 2 resume as noções básicas sobre redes credais. A seção 3 descreve o algoritmo A/R. A seção 4 apresenta o algoritmo A/R+. Adicionamente, resultados obtidos em testes preliminares, que comparam os dois algoritmos, são apresentados e discutidos.

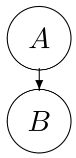
2. Redes credais

Dada uma variável aleatória X, um conjunto credal $K(X)$ é um conjunto convexo e fechado de distribuições de probabilidade sobre X. De um ponto de vista geométrico, este conjunto pode ser visto como um politopo contido num espaço do $\mathbf{R}^n$, onde n é o número de categorias de X. Desta forma, um conjunto credal pode ser representado pela enumeração das distribuições de probabilidade que definem os vértices do politopo, isto é, $K(X) = cc[p(X)_0\,, p(X)_1\,, ..., p(X)_{r-1}]$, onde r é o número de vértices e $cc[\cdot]$ indica a operação de casco convexo [H.Edelsbrunner, 1987].

Dada uma variável aleatória Y, um conjunto credal condicional $K(Y|X = x)$ tem como elementos distribuições condicionais da forma $p(Y|X = x)$. Este conjunto pode ser obtido a partir de um conjunto credal de distribuições conjuntas de probabilidade $K(XY)$. Para isto, basta aplicar a regra de Bayes a cada distribuição conjunta em $K(XY)$. O conjunto $\mathbf{Q}(Y|X) = \{K(Y|X = x_0)\,, ..., K(Y|X = x_{n-1})\}$ denota todos os conjuntos credais que podem ser obtidos pelo condicionamento de Y por X.

Dados $\mathbf{Q}(Y|X)$ e $K(X)$, é possível associar os mesmos a diferentes conjuntos credais de distribuições conjuntas. Isto porque o cômputo destes conjuntos, denominados *extensões* de $Q(Y|X)$ e $K(X)$, depende do conceito de independência condicional que é considerado existir entre $\mathbf{Q}(Y|X)$ e $K(X)$. Neste artigo é utilizado o conceito de independência forte, este conceito estabelece que duas variáveis X e Y são independentes se e somente se cada vértice em $K(XY)$ satisfaz a noção de independência estocástica. As extensões que obedecem esta definição são chamadas de *extensões fortes*.

O cômputo de uma extensão forte é extremamente complexo em termos de tempo de computação. Um algoritmo exaustivo para esta tarefa pode ser resumido da seguinte forma. Para cada maneira de combinar os vértices de $Q(Y|X)$ e $K(X)$ calcula-se uma



$$Q(B) = \{K(A) = cc[(0,5;0,5),(0,3;0,7)]\ \}$$
$$Q(B) = (K(B|a_0)\,,K(B|a_1))$$
com
$$K(B|a_0) = cc[(0,5;0,5),(0,3;0,7)]$$
$$K(B|a_1) = cc[(0,4;0,6),(0,2;0,8)]$$

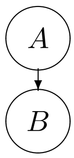
Figura 1: Exemplo de uma rede credal com variáveis binárias.

distribuição conjunta que é inserida em uma lista. A extensão forte, $K(XY)$, é o casco convexo dos pontos desta lista.

O formalismo das redes credais estabelece um esquema para representação de conhecimento incerto e impreciso que combina grafos acíclicos direcionados e conjuntos credais. Uma rede credal é definida pela dupla $\mathbf{C} = (\mathbf{G}, \mathbf{Q})$, onde $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ é um grafo acíclico e direcionado definido de maneira análoga ao de uma rede Bayesiana. Contudo, diferente de uma rede Bayesiana, cada vértice V_i de uma rede credal armazena uma coleção de conjuntos credais, denotada por $Q(V_i|pa(V_i))$, que enumera um conjunto credal para cada instanciação conjunta de $pa(V_i)$. Isto é, $Q(V_i|pa(V_i)) = (K(V_i|pa(V_i)_0)\,, ..., K(V_i|pa(V_i)_{t-1}))$, onde t é o número de realizações conjuntas em $pa(V_i)$. Se V_i é um nó raiz então seu conjunto credal não é condicional. Dado isto, o conjunto $\mathbf{Q}$ da dupla que define uma rede credal $\mathbf{C}$ contém $\{Q(V_0|pa(V_0)), ..., Q(V_m|pa(V_m))\}$, como seus elementos. Uma rede estruturada desta maneira é denominada *rede credal com conjuntos credais especificados em separado*. A Figura 1 mostra um exemplo de rede credal e seus respectivos conjuntos credais em separado.

Seja $V_q \in \mathbf{V}$, uma inferência sobre esta variável determina o limite superior e o limite inferior para a probabilidade de cada categoria de V_q. Tais limites são chamados *probabilidade inferior*, $\underline{p}(V_q = v) = \min_{p(V_q) \epsilon K(V_q)} p(V_q = v)$ e *probabilidade superior*, $\overline{p}(V_q = v) = \max_{p(V_q) \epsilon K(V_q)} p(V_q = v)$. É importante observar que a execução de inferências pode ser realizada manipulando apenas os conjuntos credais das variáveis que são relevantes para a determinação destes intervalos. Isto é, as variáveis que tem influência condicional sobre V_q. Uma característica interessante quando se aplica o conceito de independência forte sobre uma rede credal é que é possivel utilizar, como em redes Bayesianas, o critério da *d-separação* [J.Pearl, 1988] para identificar as variáveis que são relevantes para inferência [F.G.Cozman, 2000b].

Os algoritmos para inferência em redes credais podem ser classificados como sendo exatos ou aproximados. Os algoritmos exatos determinam os valores corretos para $\underline{p}(V_q = v)$ e $\overline{p}(V_q = v)$. Alguns algoritmos exatos para redes multiconectadas podem ser encontrados em [F.G.Cozman, 2000a], que apresenta um algoritmo exaustivo e [J.C.F.Rocha and F.G.Cozman, 2002], que descreve um algoritmo de eliminação de variáveis que explora a especificação em separado dos conjuntos credais e operações de casco convexo para evitar computações redundantes. [E.Fagioulі and M.Zaffalon, 1998] descreve o algoritmo 2U para inferências em polytrees com variáveis binárias.



$p(a_0) \in [0,3;0,5]$
$p(a_1) \in [0,5;0,7]$
$p(b_0|a_0) \in [0,3;0,5]$
$p(b_1|a_0) \in [0,5;0,7]$
$p(b_0|a_1) \in [0,2;0,4]$
$p(b_1|a_1) \in [0,6;0,8]$

Figura 2: Rede Bayesiana com intervalos de probabilidade.

Contudo, inferência exata em redes credais é um problema NP-Completo [J.C.F.Rocha and F.G.Cozman, 2002] e mesmo redes simples pode ser um grande desafio para algoritmos exatos. No caso da independência forte o cômputo da extensão por um algoritmo exaustivo implica na aplicação da expressão (1) para todas as combinações possíveis dos vértices pertencentes aos conjuntos credais das variáveis envolvidas na inferência. Assim, a obtenção da extensão forte de numa rede com uma topologia da forma ($X \rightarrow Y \leftarrow Z$), com variáveis quaternárias e quatro vértices por conjunto credal através de métodos exaustivos, implica no cálculo da expressão (1) sobre as 2^{36} possíveis combinações de vértices. Isto tem despertado o interesse no desenvolvimento de algoritmos aproximados.

Os algoritmos aproximados podem ser classificados com relação ao tipo de aproximação que fornecem. Uma aproximação é *externa* se suas estimativas $\underline{p}'(V_q = v)$ e $\overline{p}'(V_q = v)$ para $\underline{p}(V_q = v)$ e $\overline{p}(V_q = v)$, respectivamente, são tais que $\underline{p}'(V_q = v) \leq \underline{p}(V_q = v) \leq \overline{p}(V_q = v) \leq \overline{p}'(V_q = v)$. Os algoritmos descritos em [B.Tessem, 1992], [A.Cano and S.Moral, 2000], [V.A.Ha et al., 1998] calculam aproximações externas. Algoritmos que determinam aproximações internas ao intervalo exato são descritos em [A.Cano and S.Moral, 1996] e [A.Cano et al., 1994].

3. A/R:Inferência aproximada com intervalos de probabilidade

Em [B.Tessem, 1992] é apresentado o algoritmo A/R para inferência aproximada em redes Bayesianas com intervalos de probabilidade (uma rede Bayesiana que tem intervalos de probabilidade nas entradas de suas TPCs). Este algoritmo pode ser aplicado sobre redes credais com topologia em *polytree* desde que os conjuntos credais sejam convexos e fechados. A Figura 2 mostra a rede Bayesiana com intervalos obtida a partir da rede da Figura 1.

Dada uma variável aleatória V com categorias ($v_0, v_1, ..., v_{m-1}$), um conjunto de intervalos de probabilidade sobre V é definido por inequações que são associadas as categorias de V. Isto é, um conjunto de intervalos pode ser descrito por expressões da forma: $[p(v_l)_*; p(v_l)^*] \subseteq [0;1]$ tal que, $p(v_l) \in [p(v_l)_*; p(v_l)^*]$, para cada $l = 0..(m-1)$. Para ser aplicado, o algoritmo A/R exige que os intervalos da rede sejam consistentes. Um conjunto de intervalos é consistente quando atende as restrições $p(v_i)^* + \sum_{j=0,j\neq i}^{m-1} p(v_i)_* \leq 1$ e $p(v_i)_* + \sum_{j=0,j\neq i}^{m-1} p(v_i)^* \geq 1$. Conjuntos credais geram intervalos de probabilidade que são consistentes [B.Tessem, 1992].

O algoritmo A/R adapta o método de propagação de mensagens descrito em [J.Pearl, 1986] para inferência em redes Bayesianas com topologia em *polytree*. As alterações implementadas permitem ao algoritmo A/R manipular mensagens com intervalos e computar intervalos sobre as probabilidades marginais. As expressões (2) e (3) resumem o cálculo realizado para determinação do intervalo das probabilidades marginais para uma categoria v_l de uma variável V. Os vetores $\pi_*(V) = (\pi_*(v_0), \pi_*(v_1), ...)$ e $\pi^*(V) = (\pi^*(v_0), \pi^*(v_1), ...)$ contêm os limites superiores e inferiores dos intervalos que resumem a informação probabilística proveniente dos ascendentes de V. Os vetores $\lambda_*(V) = (\lambda_*(v_0), \lambda_*(v_1), ...)$ e $\lambda^*(V) = (\lambda^*(v_0), \lambda^*(v_1), ...)$ fazem o mesmo com relação aos descendentes de V.

$$p(V = v_l)_* = \frac{\lambda_*(v_l)\pi_*(v_l)}{\lambda_*(v_l)\pi_*(v_l) + r(l, \pi(V), \lambda(V))} \tag{2}$$

$$p(V = v_l)^* = \frac{\lambda^*(v_l)\pi^*(v_l)}{\lambda^*(v_l)\pi^*(v_l) + a(l, \pi(V), \lambda(V))} \tag{3}$$

As funções $a(\cdot)$ e $r(\cdot)$ implementam dois procedimentos chamados de aniquilamento e reforço (A/R), respectivamente. Estas operações produzem aproximações externas, tal que, o intervalo computado é mais preciso do que aquele que seria obtido simplesmente por aritmética intervalar [D.L.Draper and S.Hanks, 1995]. Apesar disto, os intervalos obtidos por A/R podem ser extremamente imprecisos [A.Cano and S.Moral, 2000].

Listagem 1 - A/R - Procedimento para o cáculo de $\pi_*(V = v_0)$

- Entrada
  - $M_*(V|X_0, X_1, ...)$
  - $\pi_V(X_0), \pi_V(X_1), ...$
- Saída: $\pi_*(V_i = v_0)$

1. Crie uma função conjunta de intervalos $\pi_{*X_0,X_1,...}(\cdot)$ e $\pi^*{}_{X_0,X_1,...}(\cdot)$, a partir de $\pi_V(X_0), \pi_V(X_1), ...$ usando multiplicação intervalar [R.E.Moore, 1966].
2. Selecione todas as entradas de $M_*(V|X_0, X_1, ...)_{i_1,...,i_k,0}$ para todas as seqüências de índices $i_1, ..., i_k$ ordene-as em ordem crescente. Insira os índices de acordo com ordenação obtida em um vetor I(1:z), onde z é o número de realizações conjuntas dos pais de V.
3. Deixe $P(X_1, X_2, ...)$ ser uma distribuição conjunta sobre $X_1, X_2, ...$. Inicialmente, faça $P(X_1, X_2, ...) = \pi_{*X_0,X_1,...}(\cdot)$. Seja S a soma da massa atribuída a $P(X_1, X_2, ...)$ e *cont* = 1.
4. enquanto ($S < 1$)
   (a) Selecione os índices $i'_1, ..., i'_k$ associados a I(cont).
   (b) Atualize $P(X_1 = x_{1,i'_1}, X_2 = x_{2,i'_2}, ...)$ com $\min(\pi_{*X_1,X_1,...}(x_{1,i'_1}, x_{2,i'_2}, ...) + 1 - S, \pi^*{}_{X_1,X_1,...}(x_{1,i'_1}, x_{2,i'_2}, ...))$.
   (c) S = S +$\pi^*{}_{X_1,X_2,...}(x_{1,i'_1}, x_{2,i'_2}, ...))$ - $\pi_{*X_1,X_1,...}(x_{1,i'_1}, x_{2,i'_2}, ...)$
   (d) *cont* = *cont* + 1.
5. Use $P(X_1, X_2, ...)$ para computar
   $\pi_*(V = v_0) = \sum_{i_1,...,i_k} M_*(V|X_0, X_1, ...)_{i_1,...,i_k,0} P(X_1 = x_{1,i_1}, X_2 = x_{2,i_2}, ...)$

Neste artigo, considera-se apenas a execução de inferências sem evidências. Neste caso, o valor dos intervalos marginais depende apenas de $\pi(V)_*$ e $\pi(V)^*$. Estes vetores são obtidos a partir das mensagens $\pi_V(X_1) = (\pi_{V*}(X_1), \pi_V^*(X_1)), \pi_V(X_2) = (\pi_{V*}(X_2); \pi_V^*(X_2)), ...$, enviadas por $pa(V) = \{X_0, X_1, ...\}$. O valor de $\pi(V)_*$ e $\pi(V)^*$ também depende da TPC $M(V|X_0, X_1, ...)$ que codifica os intervalos que relacionam V e seus pais. Os limites inferiores e superiores de $M(V|X_0, X_1, ...)$ podem ser representados pelas matrizes $M_*(V|X_0, X_1, ...)$ e $M^*(V|X_0, X_1, ...)$. A Listagem 1 mostra o algoritmo para o cômputo de $\pi_*(V = v0)$, sendo que procedimentos similares são definidos com relação as mensagens $\lambda_{Y_h}(V)$ que um nó V recebe de seus descentendes Y_h, onde $h \in \{0, 1...\}$.

4. O Algoritmo A/R+

Esta seção descreve o A/R+. Este algoritmo computa aproximações externas para inferências sobre redes credais. O algoritmo A/R+ é implementado a partir do algoritmo A/R, entretanto obtém estimativas mais precisas. Isto ocorre porque A/R+ substitui algumas operações de aniquilação e reforço por uma busca exaustiva em funções locais. A Listagem 2 descreve o procedimento usado em A/R+ para calcular sua aproximação $\breve{\pi}(V = v_0)$ para $\underline{\pi}(V = v_0)$. A aproximação que A/R+ calcula para $\overline{\pi}(V = v_0)$ é denotada por $\hat{\pi}(V = v_0)$.

Listagem 2 - A/R+ - Procedimento para o cálculo de $\breve{\pi}(V = v_0)$

- Entrada
  - $M_*(V|X_0, X_1, ...)$
  - as mensagens, vetores de intervalos de probabilidade, que V recebe de seus nós pais, $\pi_V(X_0), \pi_V(X_1), ...$,
- Saída: $\breve{\pi}(V = v_0)$

1. Para cada $\pi_V(X_0), \pi_V(X_1), ...$
   (a) Determine o maior conjunto credal que pode ser associado a mensagem;
   (b) Denote estes conjuntos credais por $K(X_0)_V$, $K(X_1)_V$, ...
2. Calcule a extensão forte destes conjuntos, $K(X_0, X_1, ...)_V$.
3. Determine o mínimo, $\breve{\pi}(V = v_0)$ a partir de
   $\sum_{i_1,...,i_k} M_*(V|X_0, X_1, ...)_{i_1,...,i_k,0} P(X_1 = x_{1,i_1}, X_2 = x_{2,i_2}, ...)$
   para todo $P \in K(X_0, X_1, ...)_V$.

Como pode ser visto, o algoritmo A/R+ utiliza uma estratégia exaustiva para calcular $\breve{\pi}(V)$. O passo 1 converte as mensagens baseadas em intervalos em conjuntos credais. O passo 2, então, computa a extensão forte $K(X_0, X_1, ...)_V$ destes conjuntos locais. Esta extensão contém a extensão exata para a mensagem conjunta que V recebe de seus pais. O passo 3 determina o mínimo da somatória $\sum M_*(V|X_0, X_1, ...)P$, para cada vértice P da extensão $K(X_0, X_1, ...)_V$. As Proposições (1) e (2) demonstram que as aproximações obtidas por A/R+ são internas às aproximações do algoritmo A/R e externas em relação aos valores exatos.

Proposição 1 *Os intervalos [$\breve{\pi}$, $\hat{\pi}$] calculados por A/R+ estão contidos (propriamente ou impropriamente), nos intervalos [π_*,π^*] obtidos por A/R.*

Tabela 1: Erro relativo na inferência $\underline{p}(E = e_0)$

Categorias por variável	N° de Vértices em cada cjto Credal	Erro Relativo A/R - Média	Erro Relativo A/R+ - Média	Amostra N° de redes
03	02	0.38	0.04	30
03	03	0.15	0.01	10
04	02	0.27	0.04	10

Prova. Em A/R+, cada mensagem $\pi_{X_k}(V)$ que um nó V recebe de seus pais é gerada recursivamente a partir de uma coleção de mensagens usadas no algoritmo A/R. Estas mensagens são usadas para determinar $K(X_0, X_1, ...)_V$, que é uma representação mais restrita do que aquela resultante do produto intervalar executado no algoritmo A/R. Logo, a somatória de produtos computada por A/R+ no passo 3 produz intervalos que são internos ou concordam com aqueles obtidos por A/R.CQD

Proposição 2 *Os intervalos [$\breve{\pi}$, $\hat{\pi}$] calculados por A/R+ contêm (propriamente ou impropriamente), os intervalos exatos [$\underline{\pi}$,$\overline{\pi}$].*

Prova. Os conjuntos credais calculados por A/R+ a partir das mensagens que um nó recebe de seus pais contêm os conjuntos credais manipulados por algoritmos exaustivos, uma vez que A/R+ utiliza o maior conjunto credal que concorda com os intervalos. Além disso, as TPCs manipuladas por algoritmos exatos possuem valores de probabilidades e não apenas seus limites. Logo, os intervalos obtidos pela somatória realizada por A/R+ são exteriores ou concordam com aqueles calculados por algoritmos exatos.CQD

Usualmente, as aproximações calculadas por A/R+ são ignificantemente mais precisas do que aquelas calculadas por A/R. Isto foi verificado através de experimentos. A Figura 3 mostra a topologia da rede credal que foi utilizada no primeiro conjunto de testes. Os conjuntos credais dessas redes foram gerados de acordo com o algoritmo dado por [J.S.Ide and F.G.Cozman, 2002].

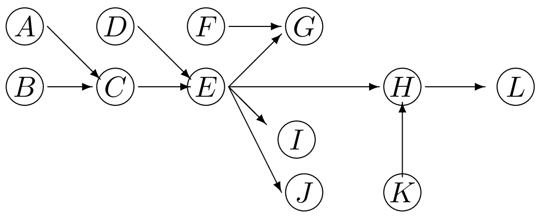


Figura 3: A topologia das redes credais usadas nos experimentos.

O média do erro relativo desses algoritmos em relação a um algoritmo exato [F.G.Cozman, 2000a], considerando uma consulta sobre a variável E em redes com variáveis ternárias e quaternárias é listada na Tabela 1. Como pode ser visto naquela tabela, o algoritmo A/R+, efetivamente, produz intervalos mais precisos do que A/R, independente da complexidade do processo de inferência. Isto também pode ser observado na Tabela 2, na qual são listadas as médias da amplitude dos intervalos resultantes de inferências sobre a variável L. Naquela tabela, como na primeira, verifica-se que os intervalos calculados por A/R+ são significativamene menores. É importante observar que da

Tabela 2: Amplitude dos intervalos para $\underline{p}(L = l_0)$, 100 amostras

N° de categorias por variável	N° de Vértices em cada cjto Credal	Amplitude A/R Média	Amplitude A/R+ Média
03	02	0,47	0,39
03	03	0,61	0,56
03	04	0,67	0,63
04	02	0,45	0,36
04	03	0,52	0,47

mesma forma que no algoritmo A/R, as mensagens que atravessam a rede no algoritmo A/R+ também são intervalos. Isto faz com que as aproximações tornem-se menos precisas a medida que a variável selecionada exige a execução de mais etapas do propagação [B.Tessem, 1992].

Um segundo conjunto de testes foi realizado sobre *polytrees* com nove diferentes topologias (cada grafo contendo 20 variáveis). Estas topologias foram geradas pelo algoritmo descrito em [J.S.Ide and F.G.Cozman, 2002] e partir de cada uma delas foram geradas aleatóriamente 5 redes credais. A seguir foram realizadas inferências sobre todas as variáveis destas redes e a amplitude dos intervalos obtidos por A/R e A/R+ foi comparada. Assim, como no caso anterior os intervalos obtidos por A/R+ (média 0,498) são menores do que aqueles calculados pelo algoritmo A/R (média 0,527). Este resultado indica que o comportamento apresentado por A/R+ é independente da topologia da rede.

Finalmente, é útil fazer uma avaliação da complexidade do algoritmo A/R+. Como A/R+ demanda o cômputo de extensões sua performance de tempo é inferior àquela de A/R. Contudo, estas extensões são locais e sua computação é, usualmente, muito mais simples do que aquela executa por algoritmos exaustivos exatos.

5. Considerações Finais

Este artigo apresentou o algoritmo A/R+, uma modificação do algoritmo A/R para inferência aproximada de intervalos probabilidade em redes credais. O algoritmo proposto produz resultados mais precisos do que o algoritmo A/R. A complexidade do algoritmo dependende apenas das estruturas locais da rede credal, o que torna o processo, usualmente, muito mais rápido do que aqueles executados por algoritmos exatos.

Futuros trabalhos objetivam determinar limitações práticas para o uso deste algoritmo dada a complexidade das redes. Adicionalmente, objetiva-se avaliar estratégias de computação que permitam combinar A/R e A/R+ na execução de inferências complexas.

Referências

A.Cano, J.E.Cano, and S.Moral (1994). Convex sets of probabilities propagation simulated annealing. In *Proc. of the 5th IPMU-94*: 978–983.

A.Cano and S.Moral (1996). A genetic algorithm to approximate convex sets of probabilities. *7th Int. Conf. IPMU-96*:847–852.

A.Cano and S.Moral (2000). Using probabilities trees to compute marginals with imprecise probabilities. Technical Report TR-DECSAI-00-02-14, University of Granada, Spain.

B.Tessem (1992). Interval probability propagation. *Int.Journal of Approximate Reasoning*, (7):95–120.

D.Heckerman, A.Mandani, and M.P.Wellman (1995). Real-world applications of Bayesian networks. *Communications of the ACM*, 38(3):24–26.

D.L.Draper and S.Hanks (1995). Localized partial evaluation of belief networks. In *11th UAI Conference*:170–177. Morgan Kaufmann.

E.Charniak (1991). Bayesian networks without tears. *AI Magazine*:50–63.

E.Fagiouli and M.Zaffalon (1998). 2U: an exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, 106(1):77–107.

F.G.Cozman (1997). Robustness analysis of Bayesian networks with local convex sets of distributions. In *13th UAI Conference*. Morgan Kaufmann.

F.G.Cozman (1999). A brief introduction to the theory of sets of probability measures. Technical report, CMU.

F.G.Cozman (2000a). Credal networks. *Artificial Intelligence*, 120(2):199–233.

F.G.Cozman (2000b). Separation properties of sets of probabilities. In *16th UAI Conference*:107–115. Morgan Kaufmann.

H.Edelsbrunner (1987). *Algorithms in Computational Geometry*. Springer-Verlag.

I.Couso, S.Moral, and P.Walley (2000). A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy*, (5):165–185.

I.Levi (1980). *The Enterprise of Knowledge*. MIT Press, 1st edition.

J.C.F.Rocha and F.G.Cozman (2002). Inference with separately specified sets of probabilities in credal networks. In *18th UAI Conference*:430–437. Morgan Kaufmann.

J.Pearl (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29(3):241–288.

J.Pearl (1988). *Intelligent Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1st edition.

J.P.Krause (1998). Learning probabilistic networks. Technical report. http://citeseer.nj.nec.com/krause98learning.html.

J.S.Ide and F.G.Cozman (2002). Random generation of Bayesian networks. In *XVI Braziliam Symposium on Artifical Intelligence*:366–375. Springer-Verlag.

R.E.Moore (1966). *Interval Analysis*. Prentice-Hall.

V.A.Ha, A.Doan, V.Vu, and P.Hadaway (1998). Geometric foundations for interval-based probabilities. In *Annals of Mathematics and Artificial Intelligence*, volume 5:582–593.