

# Reconhecimento Automático de Padrões Utilizando Redes Bayesianas: Uma Abordagem Paralela e Distribuída

Nelson Lopes Duarte Filho, Rodrigo Neves Calheiros (PIBIC/CNPq)  
Fundação Universidade Federal do Rio Grande – FURG  
Departamento de Matemática  
dmtndf@furg.br, rcalheiros@ecomf.furg.br

## Resumo

*Reconhecimento automático de padrões é um problema que encontra inúmeras aplicações, como, por exemplo, em robótica. Visando a solução desse problema, e considerando os aspectos de incerteza a ele inerentes, descreve-se uma solução que utiliza aspectos da teoria das probabilidades, em especial redes bayesianas. São apresentados algoritmos, que, utilizando técnicas de processamento paralelo e distribuído, se propõem a resolver subproblemas relacionados à aquisição automática de conhecimento (como gerar automaticamente uma rede bayesiana a partir de um banco de casos exemplo) e ao raciocínio probabilístico (como reconhecer um novo padrão apresentado, à luz do conhecimento representado numa rede bayesiana). Relacionam-se alguns resultados sobre a eficiência, obtidos numa implementação sobre uma network of workstations (NOW), em termos de speedup, e sugerem-se melhorias a serem incorporadas às soluções propostas.*

## 1. Introdução

Reconhecimento automático de padrões é um problema fundamental em inúmeras aplicações. Em robótica, por exemplo, objetos a serem manipulados e/ou trajetórias a serem percorridas são tarefas típicas que exigem identificação de imagens (situações), o que pode ser realizado com o auxílio de técnicas de reconhecimento automático de padrões. Tal problema vem sendo tradicionalmente solucionado com o uso de redes neurais. Porém, considerando tratar-se de problema envolvendo eminentemente a incerteza, propõe-se aqui o uso da teoria das probabilidades para a sua solução. Grosso modo, é admitido que, a partir da representação dos componentes de situações conhecidas, através de uma distribuição de probabilidades, reconhecer um novo padrão corresponde a computar a probabilidade dele corresponder àquelas situações, a partir da identificação dos componentes

que o constituem, e decidir entre as mais prováveis.

O seguinte exemplo hipotético simplificado, construído exclusivamente para auxiliar no entendimento da abordagem aqui proposta, parece esclarecer o problema a ser resolvido: Considere que a situação a ser identificada constitui uma imagem inscrita num quadrado (campo de visão) e pode corresponder à apenas um entre três objetos ( $obj_1$ ,  $obj_2$  e  $obj_3$ ). Ao discretizar-se o campo de visão sob a forma de um reticulado podem ser computados quantos retículos encontram-se totalmente preenchidos, semi-preenchidos e não preenchidos. Caso exista uma distribuição de probabilidades que relacione a percentagem de retículos de cada categoria com as imagens conhecidas, reconhecer o novo padrão corresponde a obter, a partir dessa distribuição, qual a probabilidade dele ser o  $obj_1$ , o  $obj_2$  ou o  $obj_3$ , dadas as percentagens de cada uma das categorias de reticulados que ele apresenta.

Identifica-se, à luz dos objetivos deste trabalho, dois subproblemas a serem resolvidos: como obter uma distribuição de probabilidades capaz de representar padrões genéricos, a partir dos seus componentes, e como, com base nessa distribuição, identificar padrões específicos. As soluções aqui propostas utilizam redes bayesianas para representar distribuições de probabilidades e se valem de dois algoritmos para atingir os objetivos vislumbrados:

- Um algoritmo capaz de construir uma distribuição de probabilidades que codifique conhecimento sob a forma de uma rede bayesiana, a partir de um banco de casos exemplo;
- Um algoritmo que utilizando essa rede seja capaz de encontrar as hipóteses que mais provavelmente suportam evidências observadas numa nova situação a diagnosticar.

Utilizando terminologia da inteligência artificial, o primeiro algoritmo costuma ser abordado no âmbito da aquisição automática de conhecimento e o segundo no do raciocínio probabilístico [1][2].

Os algoritmos que foram aqui utilizados baseiam-se em heurísticas, uma vez que os problemas por eles tratados são NP-difíceis [3]. Ainda assim, um deles é de ordem quadrática e o outro possui complexidade que cresce com a precisão exigida dos resultados, chegando a ser exponencial quando a precisão necessária for a atingível por modelos analíticos. Como ambos processam um elevado número de elementos, optou-se por implementá-los sob a forma de programas paralelos e distribuídos. Considerando que eles foram anteriormente implementados utilizando o paradigma seqüencial [4], esta nova abordagem presta-se, também, para uma comparação de resultados.

A seguir, formaliza-se o problema, descrevem-se os algoritmos utilizados para a sua solução, e como eles foram implementados sob a forma de programas paralelos e distribuídos sobre uma *NOW*. Relacionam-se alguns resultados sobre a eficiência obtida, em termos de *speedup*, e apresentam-se conclusões e trabalhos futuros a serem realizados.

## 2. Redes Bayesianas

Considerando-se que redes bayesianas são grafos dirigidos e acíclicos que representam distribuições de probabilidades, valendo-se da regra da cadeia [2]. Ou seja, partindo da hipótese que uma distribuição de probabilidades de  $n$  variáveis aleatórias  $X_1, X_2, \dots, X_{n-1}, X_n$ , pode ser representada, segundo a regra da cadeia, de acordo com a expressão:

$$P(X_1, \dots, X_n) = P(X_1) \cdot P(X_2|X_1) \cdot \dots \cdot P(X_n|X_1, \dots, X_{n-1})$$

e percebendo que, devido ao conceito de independência condicional probabilística, se  $X_i$  e  $X_k$  são independentes dado  $X_j$ :

$$P(X_i|X_j, X_k) = P(X_i|X_j),$$

conclui-se que a distribuição de probabilidades de  $n$  variáveis aleatórias pode ser representada sob a forma de uma rede bayesiana, utilizando, por exemplo, o método descrito no algoritmo 1 [5].

### Algoritmo 1 - Construção de Redes Bayesianas

1. Inicialize o grafo como vazio;
2. Para  $i = 1..n$  faça
  - (a) Escolha uma variável  $X_i$ , inexistente no grafo, e introduza-a no grafo;
  - (b) Para cada uma das variáveis  $X_j$  do grafo, menos as probabilisticamente independentes de  $X_i$ , introduza um arco dirigido de  $X_j$  para  $X_i$ ;

- (c) Associe à variável  $X_i$  a distribuição de probabilidades  $P(X_i|X_k, X_l, X_m, \dots)$ , sendo  $X_k, X_l, X_m$  as variáveis pai de  $X_i$ , determinadas em (b).

Para melhor ilustrar o método, resgata-se o exemplo introduzido na seção 1, acrescentando-se as seguintes considerações:

- Supondo o campo de visão dividido em quatro quadrantes: superior esquerdo (*SE*), superior direito (*ID*), inferior esquerdo (*IE*) e inferior direito (*ID*);
- Estabelecendo que cada quadrante é categorizado de acordo com a prevalência dos retículos que o compõem: preto (*p*), maioria de retículos totalmente preenchidos, cinza (*c*), maioria de retículos semi-preenchidos, e branco (*b*), maioria de retículos não preenchidos;
- Admitindo que a variável

$$H \in \{obj_1, obj_2, obj_3\}$$

e as variáveis

$$SE, SD, IE, ID \in \{p, c, b\}$$

são suficientes para representar o conhecimento em questão,

a rede bayesiana apresentada na figura 1 poderia ser construída, após estimados os 242 parâmetros correspondentes às distribuições de probabilidade:

$$\begin{aligned} &P(H); \\ &P(SE|H); \\ &P(SD|H, SE); \\ &P(IE|H, SD, SE); \\ &P(ID|H, SD, SE, IE). \end{aligned}$$

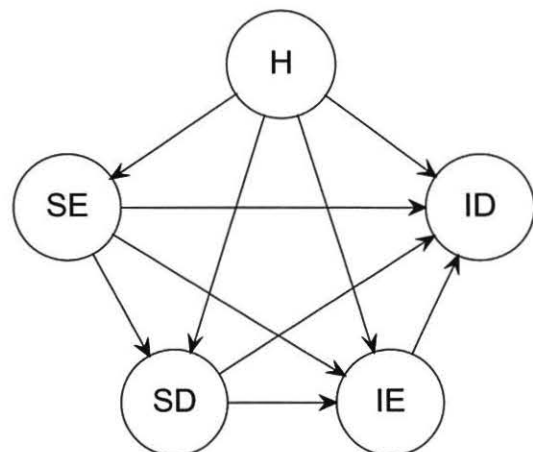


Figura 1

### 3. Aquisição Automática de Conhecimento

Para representar conhecimento, sob o ponto de vista bayesiano da teoria das probabilidades, objetivando proceder operações de raciocínio probabilístico, consideram-se as variáveis envolvidas no problema como variáveis aleatórias, e constrói-se uma rede bayesiana utilizando, por exemplo, o algoritmo 1 apresentado na seção anterior. Constitui-se, dessa maneira, um modelo computacionalmente interessante, que representa a distribuição de probabilidades conjunta de todas as variáveis de interesse [2].

Buscando uma solução para o problema do reconhecimento de caracteres, e conscientes da enorme quantidade de memória requerido para representar uma função distribuição de probabilidade de  $n$  variáveis discretas, cada uma podendo assumir  $r$  valores, Chow e Liu [6] criaram um método capaz de inferir uma distribuição de probabilidades conjunta, possível de ser representada por uma rede bayesiana com topologia em árvore, a partir das probabilidades marginais de segunda ordem das variáveis envolvidas. Sob o ponto de vista da medida de Kullback e Leibler [7], tal método obtém uma distribuição ótima com essa topologia e encontra-se descrito no algoritmo 2.

---

#### Algoritmo 2 - Construção Automática de Redes Bayesianas

---

1. Calcule as distribuições marginais de primeira e segunda ordem para todas as variáveis de interesse, a partir de um banco de casos teste;
2. Usando as distribuições do passo 1, calcule os pesos:

$$I(X_i, X_j) = \sum_{x_i, x_j} P(x_i, x_j) \log \frac{P(x_i, x_j)}{P(x_i) \cdot P(x_j)}$$

para todos os pares de variáveis e ordene-os;

3. Inicie a árvore com as duas arestas de maior peso;
  4. Examine a próxima aresta de maior peso. Se formar um ciclo descarte-a, senão conecte-a à árvore;
  5. Repita o passo 4 até que a árvore esteja concluída ( $n-1$  arestas presentes);
  6. Escolha um vértice qualquer como raiz e forme a rede bayesiana, estimando as distribuições de probabilidades condicionais de cada vértice, a partir das distribuições marginais calculadas no passo 1.
- 

Esse algoritmo, capaz de adquirir conhecimento automaticamente, foi implementado sob a forma de um programa paralelo e distribuído, utilizando o paradigma mestre-escravo: o mestre lê um banco de casos exemplo e o distribui entre os escravos. Cada um destes, por sua vez, realiza

as operações descritas nos passos 1 e 2 do algoritmo, sobre os pares de vértices a eles designados, a partir da parte do banco recebida do mestre. Cada escravo recebe aproximadamente o mesmo número de pesos a calcular. Após calculados os pesos, estes são enviados de volta ao mestre, que os utiliza para construir a árvore (passos 3 a 6 do algoritmo).

Ainda lembrando o exemplo introduzido na seção 1, a rede bayesiana em forma de árvore, mostrada na figura 2, poderia eventualmente ser obtida pelo algoritmo.

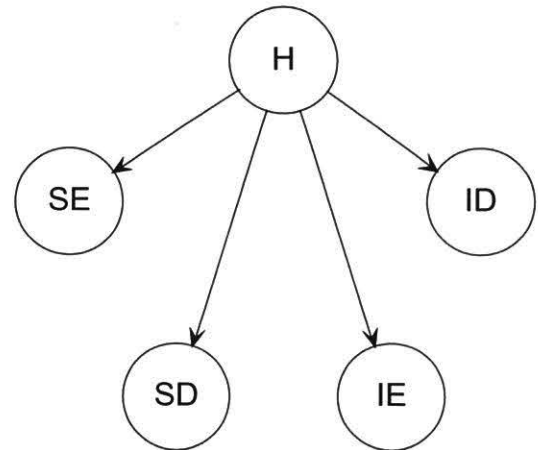


Figura 2

Para essa rede seria necessário estimar os 25 parâmetros correspondentes às distribuições de probabilidades:

$$\begin{aligned} &P(H); \\ &P(SE|H); \\ &P(SD|H); \\ &P(IE|H); \\ &P(ID|H). \end{aligned}$$

Tais estimativas poderiam ser realizadas a partir do banco de casos exemplo, utilizando as seguintes formulações típicas:

- $P(H = obj_1) = \text{Número de casos exemplo que descrevem objetos do tipo } obj_1 \text{ dividido pelo número total de casos exemplo};$
- $P(SE = p|H = obj_1) = \text{Número de casos exemplo que descrevem objetos do tipo } obj_1, \text{ cujo quadrante superior esquerdo contém maioria de retículos preenchidos, dividido pelo número de casos exemplo que descrevem objetos do tipo } obj_1;$
- e assim sucessivamente.

### 4. Raciocínio Probabilístico

O reconhecimento de padrões, no contexto aqui apresentado, corresponde a resolver o problema da incorporação de

evidências durante uma operação de raciocínio probabilístico [2] [5]. Especificamente, consiste em obter as distribuições de probabilidade condicional das variáveis consideradas hipóteses, condicionadas aos valores das variáveis consideradas evidências, a partir da distribuição de probabilidades conjunta codificada na rede bayesiana em questão. Formalmente, equivale a obter as distribuições de probabilidades  $P(X_i|x_j, x_k, x_l, \dots)$  para todo  $X_i$  que represente um conjunto possível de hipóteses, sendo  $x_j, x_k, x_l, \dots$  os valores de todas as variáveis  $X_j, X_k, X_l, \dots$  observadas (evidências). O método aqui adotado para realizar essa tarefa encontra-se mostrado no algoritmo 3 e foi proposto por Pearl [2]. Baseia-se em simulações estocásticas e resume-se em simular amostras, atribuindo valores aleatórios para as variáveis da rede, sorteados segundo as probabilidades nela representadas.

---

### Algoritmo 3 - Reconhecimento de Padrões

---

1. Obtenha uma amostra aleatória, sorteando, recursivamente, um valor para cada uma das variáveis da Rede Bayesiana, utilizando como chances os valores das probabilidades condicionais registradas nos vértices, condicionadas aos valores sorteados para as variáveis que as antecedem no grafo (para as variáveis raiz, utilize as probabilidades marginais);
  2. Caso os valores sorteados para as variáveis evidência no passo 1 coincidam com os valores observados no padrão a ser reconhecido, considere a amostra como válida, do contrário descarte-a;
  3. Se o número de amostras aleatórias válidas é insuficiente para a realização das inferências estatísticas necessárias, de acordo com precisão estabelecida, volte ao passo 1;
  4. Estime as distribuições de probabilidade das variáveis hipótese, a partir da frequência de ocorrência dos seus valores, no conjunto de amostras aleatórias válidas. Essas distribuições de probabilidade são as que dão suporte à escolha da hipótese (padrão) mais plausível.
- 

Esse algoritmo é trivialmente paralelizável [8] e foi sintetizado sob a forma de um programa implementado com base no paradigma da divisão e conquista: os processos executam paralelamente até gerar um número definido de amostras válidas.

Para a geração dessas amostras é necessário que os processos possuam uma descrição da rede a ser analisada. Um dos processos é o responsável pela obtenção dessa descrição e da sua distribuição para os demais. Estes então realizam as simulações individualmente, até que  $N$  amostras válidas sejam obtidas. O número de amostras geradas é igual para

todos os processos. Assim, para obter  $N$  amostras válidas, cada processo realiza  $N/P$  simulações válidas, sendo  $P$  o número de processos.

As probabilidades de ocorrência dos valores das variáveis hipótese são obtidos pelo número de vezes em que eles ocorreram nas amostras válidas, dividido pelo número total de amostras válidas geradas pelos processos. Para que essas probabilidades possam ser computadas, cada processo envia os seus resultados parciais para o processo responsável pela distribuição da rede.

Nota-se nessa implementação que a comunicação entre processos é bastante reduzida: somente circulam mensagens contendo a descrição da rede bayesiana e os resultados parciais das simulações realizadas pelos processos.

Ainda resgatando o exemplo introduzido na seção 1 e supondo que tenham sido identificadas as evidências:

$$\begin{aligned} SE &= p; \\ SD &= c; \\ IE &= p; \\ ID &= b, \end{aligned}$$

seriam consideradas válidas as amostras:

$$\langle H, SE, SD, IE, ID \rangle \in \{ \langle \_, p, c, p, b \rangle \}.$$

A probabilidade da imagem corresponder ao objeto  $obj_1$ , é dada pela razão entre o número de amostras que contém  $H = obj_1$  e o número total de amostras. Idem para os demais objetos.

## 5. Resultados

Objetivando avaliar o desempenho dos algoritmos paralelos de geração automática de redes bayesianas e de reconhecimento automático de padrões, foram realizadas simulações. Para isso, os algoritmos acima descritos foram implementados em linguagem C, utilizando o compilador GCC e a biblioteca LAM/MPI.

Os programas foram executados em um NOW composto por 8 Pentium II 200MHz com 32MB de RAM, conectados a uma rede Ethernet de 10Mbps, rodando o sistema operacional Debian GNU/Linux.

Os testes foram realizados utilizando variáveis aleatórias binárias, e o algoritmo de diagnóstico utilizou como critério de parada a geração de 10.000 amostras válidas. Os resultados obtidos encontram-se resumidos nas tabelas 1 e 2. Ali, são fornecidos os valores de *speedup* médio dos dois algoritmos, sendo  $n$  o número de vértices das redes bayesianas.

Apesar do algoritmo para geração automática de redes não apresentar um ganho de desempenho muito elevado, este pode ser considerado satisfatório, uma vez que o algoritmo é executado esporadicamente, apenas quando se faz necessária a atualização do conhecimento representado na

	n=50	n=100
p=1	1,000	1,000
p=4	1,666	2,332
p=8	1,489	2,346

**Tabela 1. Speedup do algoritmo de geração de redes**

	n=50	n=100	n=200
p=1	1,000	1,000	1,000
p=2	1,945	1,955	1,962
p=4	3,575	3,606	3,572
p=8	6,234	6,554	6,696

**Tabela 2. Speedup do algoritmo de reconhecimento de padrões**

rede. Também é importante salientar, apesar de não ter sido relacionado, que o ganho de desempenho desse algoritmo melhora a medida em que aumenta o número de valores que as variáveis aleatórias podem assumir.

Quanto ao algoritmo de reconhecimento automático de padrões, percebe-se um bom ganho de desempenho.

## 6. Conclusões e Trabalhos Futuros

Foram apresentados algoritmos para aquisição automática de conhecimento e para realização de raciocínio probabilístico, utilizando aspectos da teoria das probabilidades, especialmente de redes bayesianas. Esses algoritmos foram implementados sob a forma de programas paralelos e distribuídos e com eles realizadas simulações. Das simulações conclui-se que a solução apresentada para o problema de reconhecimento automático de padrões mostra-se adequada.

Destarte, pretende-se, à continuação dos trabalhos, aprimorar os algoritmos e suas implementações, de modo a atingir um desempenho ainda maior. De antemão, percebe-se que o algoritmo de geração automática de redes pode ser melhorado, se o método de distribuição de tarefas entre os escravos tiver reduzido o tempo de comunicação entre os processos.

Já com o algoritmo proposto para reconhecimento automático de padrões, apesar do bom desempenho que apresenta, pode ainda ser melhorado consideravelmente, se for adaptado para realização de consultas em redes específicas. Assim sendo, a distribuição da descrição da rede, responsável pela maior parte do tráfego de comunicação entre os processos, seria evitada.

Pretende-se, também, utilizar as soluções aqui apresentadas em situações reais, especialmente nas aplicações realizadas pelo Grupo de Pesquisa em Robótica e Inteligência

Artificial da FURG. Entende-se que essa intenção será facilitada, tendo em vista o caráter genérico dos algoritmos propostos e implementados.

## Referências

- [1] E. Rich and K. Knight, *Inteligência Artificial*, Makron Books do Brasil, Rio de Janeiro, 1994.
- [2] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference*, Morgan Kaufmann Publishers, Inc., Sao Mateo, 1988.
- [3] G. F. Cooper, *Probabilistic Inference Using Belief Networks is NP-hard*, Report KSL-87-27, Medical Computer Science Group, Stanford University, 1987.
- [4] C. Z. Billa, *Representação do Conhecimento Através de Redes Bayesianas*, Projeto de Graduação Apresentado ao Curso de Engenharia de Computação, FURG, 2001.
- [5] N. L. Duarte Filho, *Raciocínio Evidencial e Aquisição Automática de Conhecimento: Uma Abordagem Bayesiana*, Tese de Doutorado, Departamento de Informática, PUC/RJ, 1991.
- [6] C. Chow, C. Liu, *Aproximating Discrete Probability Distributions with Dependence Trees*, IEEE Transactions on Information Theory, IT-14: 462-467, 1968.
- [7] S. Kullbak, R. A. Leibler, *Information an Sufficiency*, Ann. Math. Statistics, 22:79-86, 1951.
- [8] N. Lemke, *Aplicações de Alto Desempenho Trivialmente Paralelizáveis*, Anais: 2ª Escola Regional de Alto Desempenho, editores T.A. Diverio, G.G.H Cavalheiro, Porto Alegre, 2002, pp. 107-138.