

# **Otimização do aprendizado de redes Bayesianas por meio de uma ordenação de variáveis adequada**

**Instituição:** Universidade Federal de São João del-Rei / Departamento de Ciência da Computação.

**Proponente:** Edimilson Batista dos Santos.

**Palavras Chave:** aprendizado de máquina; redes Bayesianas; aprendizado sem fim.

## **1. Introdução**

Um dos maiores desafios dentro da área de Inteligência Artificial é desenvolver sistemas que possam agir racionalmente. Para alcançar este objetivo, muitas dificuldades precisam ser superadas. Por exemplo, é preciso considerar que nem sempre há conhecimento completo sobre um domínio. Assim, um agente inteligente deverá agir sob a incerteza.

Nas últimas décadas, as Redes Bayesianas (PEARL, 1988) tornaram-se um método muito utilizado para representação do conhecimento e do raciocínio envolvendo incerteza. Este formalismo aplica teoria de grafos e teoria de probabilidade sendo formado basicamente por 2 componentes: a estrutura gráfica (grafo direcionado acíclico) e os parâmetros numéricos (tabelas de probabilidades condicionais). Sendo assim, redes Bayesianas são consideradas representações gráficas de distribuição de probabilidade e têm sido aplicadas no desenvolvimento de modernos sistemas especialistas, máquinas de diagnósticos e sistemas de suporte à decisão (HECKERMAN, 1998).

Redes Bayesianas podem ser modeladas a partir de um conjunto de dados. Este processo, conhecido como aprendizado, divide-se em duas etapas: i) aprendizado da estrutura gráfica (grafo direcionado acíclico) e ii) aprendizado dos parâmetros numéricos (distribuições das probabilidades da rede). Aprender os parâmetros numéricos pode ser considerado uma tarefa simples. Contudo, aprender a estrutura é um problema complexo, considerado NP-Completo (CHICKERING, 1996). Assim, muitos pesquisadores buscam encontrar novos métodos que possam otimizar o aprendizado da estrutura com auxílio de outras técnicas, como por exemplo algoritmos evolutivos.

Os métodos Bayesianos para o aprendizado da estrutura dividem-se em duas classes principais. A primeira é a classe dos algoritmos que geram a rede através de uma busca heurística. Já na segunda classe, estão os algoritmos que utilizam o conceito de independência condicional (NEAPOLITAN, 2003) para a construção da rede.

Um bom exemplo de algoritmo com base em busca heurística é o K2, proposto por Cooper e Herskovitz (COOPER e HERSKOVITS, 1992). Este algoritmo é muito conhecido devido ao seu desempenho em termos de complexidade computacional (tempo) e resultados precisos, obtidos quando uma ordenação adequada de variáveis é fornecida. A suposição dos atributos pré-ordenados é usada para reduzir o número de possíveis estruturas a serem aprendidas. No entanto, encontrar tal ordenação para as variáveis em uma rede Bayesiana é

um problema complexo, que exige muita informação sobre o modelo (De Campos et al., 2000).

A informação da ordenação de variáveis pode não estar disponível em aplicações do mundo real (CHEN et al., 2008) e, muitas vezes, é apresentada de forma aleatória aos algoritmos de aprendizado, o que leva a resultados pobres (HRUSCHKA Jr. et al., 2007). Uma boa ordenação das variáveis pode melhorar os resultados do aprendizado de uma rede Bayesiana, no entanto, uma ordem que não represente de forma correta o relacionamento das variáveis pode inserir erros no processo de aprendizagem. Realizar uma busca exaustiva através de todas as ordenações para problemas grandes, pode ser intratável. Assim, algumas heurísticas têm sido propostas para encontrar uma ordenação adequada.

Este trabalho foca na busca de uma ordenação de variáveis adequada para o aprendizado de redes Bayesianas, utilizando para isto técnicas de Seleção de Atributos (*Feature Selection*). Mais especificamente, pretende-se explorar duas técnicas conhecidas como Coeficiente de Correlação (HSU et al., 2010) e Boruta (KURSA e RUDNICKI, 2010), além da utilização de técnicas de Importância dos Atributos (*Feature Importance*), as quais consistem em ranquear os atributos de um conjunto de dados de acordo com a ordem de contribuição de cada um deles.

## 2. Redes Bayesianas

Redes Bayesianas são representações gráficas de distribuição de probabilidades e têm sido muito utilizadas para representação de incerteza em Inteligência Artificial. Segundo (HECKERMAN et al., 1998), tais representações possuem um papel crucial em modernos sistemas especialistas, máquinas de diagnósticos e sistemas de suporte à decisão.

Uma rede Bayesiana consiste de dois componentes importantes:

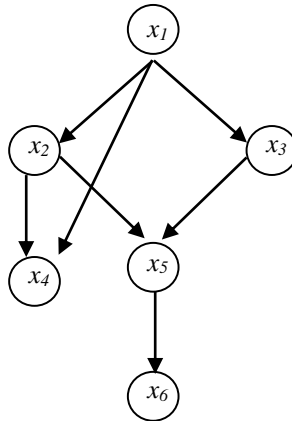
- Um grafo direcionado acíclico (DAG)  $G = (V, E)$  – ou estrutura gráfica –, onde  $V = \{X_1, X_2, \dots, X_n\}$  é o conjunto de nós e representa as variáveis aleatórias e  $E$  é o conjunto de pares ordenados de elementos distintos de  $V$  e representa as relações de dependências entre as  $n$  variáveis. Os elementos de  $E$  são chamados de arestas (ou arcos). Uma aresta direcionada do nó  $X_i$  para o nó  $X_j$  indica que  $X_i$  é um dos pais de  $X_j$ .
- Uma tabela de probabilidade condicional (CPT) – ou parâmetros numéricos –, que quantifica os efeitos que o conjunto de pais de  $X_i$  tem sobre as variáveis  $X_i$  em  $G$ .

Numa rede Bayesiana, a ausência de alguns arcos no DAG (conhecido como componente qualitativo) demonstra a existência de relações de independência condicional entre as variáveis e a presença deles pode representar a existência de relações de dependência direta. A tabela de probabilidades (conhecida como componente quantitativo) é uma coleção de medidas de probabilidades condicionais, as quais demonstram a força das dependências e são atualizadas com o uso do Teorema de Bayes, com base em uma nova informação amostral.

A Figura 1 apresenta um exemplo de rede Bayesiana e mostra a relação direta existente entre suas variáveis, ligadas por arcos orientados. Observa-se que  $x_i$  está diretamente relacionada com  $x_j$  se houver um arco de  $x_i$  para  $x_j$ . Alguns autores consideram a relação entre as variáveis como sendo uma relação causal (PEARL, 1988) e esta relação tem as seguintes possibilidades:

- Uma variável pode causar uma ou mais variáveis (filhas).

- Uma variável pode sofrer a influência causal de uma ou mais variáveis (pais).



**Figura 1. Estrutura de Rede Bayesiana.**

A quantificação da relação causal é atualizada por uma distribuição de probabilidade condicional, a qual condiciona a variável causada à(s) sua(s) causadora(s) –  $P(\text{causada}|\text{causadora})$ . Representa-se a distribuição conjunta como o produto das distribuições condicionais dadas pelas relações causais na rede. Assim, para a rede da Figura 1, tem-se a equação (1):

$$P(x_1, x_2, \dots, x_6) = P(x_6|x_5) P(x_5|x_2, x_3) P(x_4|x_1, x_2) P(x_3|x_1) P(x_2|x_1) P(x_1) = \prod_{i=1}^m P(x_i | \pi_{x_i}) \quad (1)$$

na qual  $m$  é o número de variáveis,  $P(x_1, x_2, \dots, x_6)$  é a probabilidade conjunta de  $X = \{x_1, x_2, \dots, x_6\}$ , sendo  $X$  um conjunto de variáveis aleatórias,  $P(x_i|x_j)$  é a probabilidade condicional de  $x_i$ , dado que se conhece  $x_j$ , e  $\pi_{x_i}$  é o conjunto de pais de  $x_i$ .

Num modelo causal, não há relações com complexidade muito grande e a estrutura de cálculos terá uma complexidade proporcional ao número de variáveis dependentes e não ao número de variáveis do problema, sendo esta uma das grandes vantagens em se utilizar este formalismo.

### 2.1. Algoritmo K2

K2 (COOPER e HERSKOVITS, 1992) é um algoritmo de busca heurística criado para o aprendizado de estruturas de redes Bayesianas. Ele recebe, como entrada, um conjunto de dados e uma ordenação das variáveis e gera, como saída, a estrutura da rede. K2 usa uma lista ordenada (contendo todos os atributos) que afirma que apenas os atributos posicionados antes de um dado atributo  $x_i$  podem ser pais de  $x_i$ . Assim, o primeiro atributo na lista não tem pais, ou seja, é um nó raiz na rede Bayesiana.

O algoritmo usa um método guloso para encontrar a melhor estrutura. Ele se inicia supondo que todos os nós não têm pais. Então, a partir do segundo atributo da lista ordenada (o primeiro é um nó raiz), os possíveis pais são testados e aqueles que maximizam a probabilidade da estrutura estar de acordo com a base de dados são adicionados à rede. Este processo é repetido para todos os atributos até que a melhor estrutura possível seja encontrada.

A métrica (pontuação) aplicada pelo K2 para testar cada conjunto de pais possíveis, para cada atributo, é dada por  $g(i, \pi_i)$ :

$$g(i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!$$

onde cada atributo  $x_i$  ( $i = 1, \dots, n$ ) tem  $r_i$  possíveis valores ( $v_{i1}, v_{i2}, \dots, v_{iri}$ ). Cada atributo  $x_i$  tem um conjunto de pais  $\pi_i$  e  $q_i$  é o número de instâncias de  $\pi_i$ .  $N_{ijk}$  é o número de objetos no conjunto de dados  $D$ , onde  $x_i$  tem o valor  $v_{ik}$  e  $\pi_i$  é instanciado como  $w_{ij}$ , o qual representa a  $j$ -ésima instância relativa a  $D$  de  $\pi_i$ . Finalmente,  $N_{ij} = \sum_k N_{ijk}$ .

### 3. Metodologia e Cronograma

O desenvolvimento científico do projeto será realizado através da continuidade dos trabalhos já iniciados e que vêm sendo desenvolvidos, e buscando-se atuar principalmente em pontos onde há carências específicas de melhores resultados. A metodologia planejada para este projeto pode ser resumida através dos seguintes passos metodológicos:

#### 1. Pesquisa e investigação teórica com base na literatura;

Este passo metodológico será iniciado por um aluno de iniciação científica e continuará a ser desenvolvido em paralelo com as outras atividades investigativas. O aluno de iniciação científica será orientado a pesquisar sobre redes Bayesianas e seus algoritmos de aprendizado, com foco principalmente no algoritmo K2.

Sob a orientação do proponente, o aluno também será capacitado para o entendimento e utilização da teoria de redes Bayesianas no processo de inferência.

Será realizado um levantamento bibliográfico sobre trabalhos que investiguem o uso da ordenação de variáveis para otimização do aprendizado de redes Bayesianas.

#### 2. Estudo sobre bibliotecas para redes Bayesianas;

Diversas bibliotecas têm sido implementadas em diferentes linguagens de programação para a aplicação de redes Bayesianas. Será realizado um estudo sobre estas bibliotecas e escolhida uma delas para a implementação dos algoritmos de aprendizado e execução dos experimentos.

#### 3. Estudo e implementação do algoritmo K2;

O algoritmo K2 será estudado e implementado com auxílio da biblioteca escolhida no passo 2.

#### 4. Estudo e implementação de técnicas de seleção de atributos;

Serão estudadas algumas técnicas de seleção de atributos, com destaque inicial para Coeficiente de Correlação e Boruta. Também serão estudadas técnicas de importância dos atributos, como a técnica *Random Forest Feature Importance*. Em seguida, estas técnicas serão implementadas (ou aplicadas a partir da biblioteca escolhida) para gerar ordenações de variáveis adequadas à execução do K2.

#### 5. Execução de experimentos com o K2 e ordenações adequadas;

Neste passo, o algoritmo K2 será executado com bases de dados clássicas da literatura de redes Bayesianas, recebendo também como entrada as ordenações criadas pelas técnicas de seleção de atributos e/ou importância dos atributos. Os resultados obtidos pelo K2 com as

ordenações adequadas serão comparados aos resultados obtidos pelo K2 com ordenações aleatórias, investigando se as ordenações criadas pelas técnicas de seleção de atributos são capazes de melhorar o aprendizado das redes Bayesianas.

## 6. Escrita do relatório final.

Nesta etapa, este projeto é finalizado com a escrita do relatório final, que servirá como instrumento de planejamento para a sequência da pesquisa.

A sequência temporal do desenvolvimento dos passos metodológicos definidos acima, durante os 12 meses previstos para a execução do projeto, pode ser vista na Tabela 1.

**Tabela 1. Cronograma Técnico-Científico.**

	mês1	mês2	mês3	mês4	mês5	mês6	mês7	mês8	mês9	mês10	mês11	mês12
Passo1												
Passo2												
Passo3												
Passo4												
Passo5												
Passo6												

## 4. Plano de Trabalho

O aluno de iniciação científica participará de cada uma das etapas de desenvolvimento do projeto, já apresentadas na seção 4, com supervisão e orientação do proponente. A etapa 1 levará a investigações iniciais e a um levantamento bibliográfico sobre definições e métodos de aprendizado de redes Bayesianas.

Na etapa 2, o aluno deve estudar e definir uma linguagem de programação e uma biblioteca que o ajudará na implementação dos algoritmos K2 e DMBC. Assim, na etapa 3, o aluno deverá implementar estes dois algoritmos. Na etapa 4, serão realizados os experimentos com os algoritmos implementados, aplicando-os a bases de dados de redes Bayesianas para comparação e análise dos resultados. Na etapa 5, o aluno será orientado na escrita de relatório e artigo científico.

O desenvolvimento do aluno será avaliado por critérios como assiduidade e pontualidade, iniciativa, habilidade técnica, criatividade, expressão oral e escrita, entre outros, durante todo o trabalho. O resultado dessa avaliação será aplicado constantemente no aperfeiçoamento das habilidades do aluno.

## Referências

Chen, X.; Anantha, G.; Lin, X. Improving Bayesian Network Structure Learning with Mutual Information-based Node Ordering in the K2 Algorithm. IEEE Transactions on Knowledge and Data Engineering, vol. 20, n 5, p. 628-640, 2008.

Chickering, D.M. Learning Bayesian networks is NP-Complete. In: FISHER, D.; LENZ, H (Eds). Learning from Data: artificial intelligence and statistics V. Springer-Verlag, p. 121-130, 1996.

Cooper, G. F.; Herskovits, E. 1992. A Bayesian Method for the Induction of Probabilistic Networks from Data. *Mach. Learn.* 9, 4 (October 1992), 309-347.

De Campos, L. M.; Huete, J. F. Approximating causal orderings for Bayesian networks using genetic algorithms and simulated annealing. In: Proceedings of the Eighth International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU'00), v. 1, p. 333–340, 2000.

Heckerman, D. A tutorial on learning Bayesian networks. Local: Microsoft Research, Advanced Technology Division, 1995, Notas: Technical Report MSRTR-95-06.

Hruschka Jr., E. R.; Ebecken, N. F. F. Towards efficient variables ordering for Bayesian Network Classifiers optimization. *Data & Knowledge Engineering*, v. 63, p. 258-269, 2007.

Hsu, Hui-Huang et al. Feature Selection via Correlation Coefficient Clustering. In *J. Softw.*, v. 5, n. 12, p. 1371-1377, 2010.

Kursa, M. B.; Rudnicki, W. R. Feature Selection with the Boruta Package. In *Journal of Statistical Software*, [S. l.], v. 36, n. 11, p. 1–13, 2010. DOI: 10.18637/jss.v036.i11.

Neapolitan, R. E. Learning Bayesian networks. New Jersey: Prentice Hall, Inc., 2003.

Pearl, J. Probabilistic reasoning in intelligent systems: networks of plausible inference. San Mateo: Morgan Kaufmann, 1988.

Santos, E. B.; Hruschka Jr., E. R.; Hruschka, E. R.; Ebecken, N. F. F., Bayesian network classifiers: Beyond classification accuracy. *Intelligent Data Analysis (IDA)*, v. 15, n. 3, p. 279-298, 2011.