Proceedings of the VI Brazilian Conference on Neural Networks - VI Congresso Brasileiro de Redes Neurais
pp. 187–192, June 2–5, 2003 - Centro Universitario da FEI, Sao Paulo - SP - Brazil

Um Critério para Retirada de Arcos em Redes Bayesianas

Luiz G. Q. S. Junior[1], E. C. Gurjão[1], F. M. de Assis[1], Ernesto L. Pinto[2]
[1]Departamento de Engenharia Elétrica,
Universidade Federal de Campina Grande, 58100-000,
Campina Grande, PB. Brasil.
[2] Departamento de Engenharia Elétrica,
Instituto Militar de Engenharia,
Rio de Janeiro, Brasil.
E-mails: {junior,ecandeia,fmarcos}@dee.ufcg.edu.br, ernesto@aquarius.ime.eb.br

Abstract

In Bayesian networks propagation of information is done with messages sent between adjacent nodes. However, if some nodes of a network are in loop, messages can circulate endlessly between them, giving rise to a serious setback concerning the application of such network. Some techniques have been proposed to overcome this problem. One of them consists of breaking the loop by appropriate removal of one of its arcs. In this paper a new criterion is proposed for selecting arcs to be removed with the same aim. This criterion is based on the conditioned entropy of random variables assigned to the nodes that form an arc. An illustrative example of its application is also presented.

1. Introdução

O diagnóstico médico se insere numa categoria ampla de problemas nos quais a questão central é inferir sobre a ocorrência de uma ou mais anomalias internas a um sistema a partir da observação de indicadores externos por ele produzidos.

Fornecer um diagnóstico pode ser uma tarefa difícil em alguns casos devido a incertezas nos sintomas e à possibilidade de um mesmo sintoma estar associado a mais de uma doença. No caso específico das doenças neuromusculares ainda há uma outra característica peculiar: não é desprezível a possibilidade de duas anomalias (doenças) afetarem as observações (sintomas) de maneiras opostas. Todas estas características evidenciam a complexidade do diagnóstico médico e a utilidade de se dispor de ferramentas computacionais para o auxílio ao trabalho do médico.

Um problema subjacente ao desenvolvimento destas ferramentas é a modelagem da relação probabilística entre o conjunto de anomalias de interesse (doenças) e o conjunto de observações de que se pode dispor (indicadores de exames laboratoriais e sintomas observados em exames clínicos).

Dadas as peculiaridades desta relação, é especialmente importante que a modelagem viabilize a incorporação do conhecimento empírico acumulado pelos médicos no desenvolvimento e aplicação das ferramentas de auxílio ao diagnóstico.

Nesse sentido, as redes probabilísticas se mostram adequadas. Essas redes podem ser entendidas de maneira simples: são grafos orientados onde as ligações entre dois nós são munidas de medidas de probabilidade condicional e que podem ser construídos a partir de uma base de dados empíricos (observações e os diagnósticos correspondentes).

A idéia básica que alimenta o grande interesse pelas redes probabilísticas, também conhecidas como redes bayesianas, é fatorar uma função distribuição de probabilidade com múltiplas variáveis em diversos termos mais simples (envolvendo poucas variáveis), explorando relações de independência condicional entre as variáveis aleatórias em questão.

Uma vez obtida a rede bayesiana associada a uma base de dados, é possível usá-la para avaliar o impacto de novas entradas no sistema modelado e realizar inferências de base probabilística.

No projeto "Sistema Computacional para Auxílio ao Diagnóstico de Doenças Neuromusculares" [1], que deu origem a este trabalho, foi proposto o uso das redes bayesianas para o desenvolvimento de uma ferramenta computacional de auxílio ao diagnóstico das doenças nele contempladas. Devido a problemas encontrados nos algoritmos de inferência quando a rede apresenta laços, neste artigo propomos um procedimento para se obter uma rede derivada sem laços, através da retirada criteriosa de arcos.

O trabalho apresenta na Seção 2 as técnicas de construção de uma rede a partir dos dados. Na Seção 3 apresenta-se as técnicas para inferência e os problemas com a presença de laços. Na Seção 4 é apresentado o critério para a retirada dos laços. Na Seção 5 esse critério é aplicado à rede obtida em [1] e finalmente na Seção 6 são apresentadas as conclusões e tecidos alguns comentários sobre possíveis extensões do presente trabalho.

2. Construção e Treinamento de uma Rede Probabilística

Discute-se nesta seção a montagem de uma rede probabilística associada a uma base de dados definida a seguir.

Definição 1 *Seja U um conjunto de variáveis aleatórias discretas $\{X_1, ..., X_n\}$, $n \geq 1$. Um* **caso** *em U consiste numa atribuição de valores para todos os $X_i \in U$. Uma* **base de dados de casos** *D sobre U é uma lista de casos. Considera-se que os casos na base de dados são independentes, e assim não importa a ordem em que eles ocorrem.*

Partindo de uma base de dados D , o problema que se coloca é construir a estrutura de rede probabilística na qual as relações entre essas variáveis estarão representadas por arcos que levam de pais para filhos com as respectivas distribuições de probabilidades condicionais. Esta tarefa pode ser desmembrada naturalmente em duas: "aprender" a estrutura da rede B_S (a topologia que estabelece as relações entre as v.a. X_i) e estimar um conjunto de probabilidades condicionais B_P a ela associada.

Em princípio, a estrutura da rede poderia ser obtida dos dados por um procedimento do tipo "tentativa e erro" onde todas as possíveis combinações de nós seriam testadas no sentido de se avaliar "quem pode ser pai de quem", com base nas estimativas das distribuições de probabilidades condicionais correspondentes. No entanto, devido ao grande número de combinações, o teste de todas as possibilidades é impraticável e para tanto deve ser usado um algoritmo que elimine parte dessas possibilidades.

Os algoritmos propostos com este objetivo podem ser separados em duas categorias: bayesianos e não bayesianos. Estes últimos usam testes estatísticos na base de dados para decidir a existência de arcos na rede a ser construída.

Por outro lado, os algoritmos bayesianos assumem a existência de uma distribuição de probabilidades *a priori* sobre todas as possíveis redes, calculam distribuições atualizadas ("a posteriori") usando a base de dados disponível e escolhem a rede com nelhor distribuição atualizada. Uma questão central nestes métodos bayesianos é o estabelecimento de um critério de qualidade adequado para avaliação das distribuições atualizadas das estruturas de rede.

A abordagem bayesiana apresenta algumas vantagens bem conhecidas. Pelo fato de não usar testes estatísticos de dependência condicional entre variáveis estes métodos são comparativamente mais adequados do que os métodos não bayesianos para o emprego com bases de dados de tamanho limitado. Por outro lado, a abordagem bayesiana fornece critérios de parada naturais para os algoritmos, em lugar de comparações com limiares estabelecidos de forma arbitrária, que é usual em métodos não bayesianos. Por fim, a abordagem bayesiana permite incorporar de maneira muito fácil o conhecimento "a priori" sobre o domínio de interesse.

O algoritmo bayesiano denominado K3 tem sido objeto de diversos trabalhos publicados nos últimos anos [2] e usa um critério de qualidade baseado no princípio da descrição pelo comprimento mínimo (MDL, do inglês *minimum description length*), princípio este originalmente aplicado ao problema de codificação de fontes de informação [3]. A idéia básica do algoritmo é buscar a estrutura de rede probabilística que possibilitaria uma representação da base de dados com a menor quantidade de símbolos possível.

O procedimento proposto em [2] parte de uma ordenação prévia das variáveis e de uma rede constituída de somente um nó, representando a primeira variável. A cada passo do algoritmo um novo nó é inserido e todos os nós incluídos previamente são avaliados como candidatos a serem pai do novo nó. Para cada possibilidade é avaliado uma medida associada ao critério MDL $L(B_S, D)$, que será aqui denominada *métrica de descrição* ou *métrica DL* e é função da rede B_S e da base de dados D. A alternativa (associação pai-filho) com menor valor da métrica de descrição é escolhida para compor atualização da rede em construção. O procedimento prossegue até que todas as variáveis sejam incluídas na estrutura de rede.

Uma vez obtida a estrutura B_S é necessário voltar ao conjunto de dados para estimar as probabilidades condicionais associadas a esta rede, ou seja o conjunto B_P. Uma alternativa para isto é o procedimento proposto em [4], que foi empregado no projeto que deu origem ao presente artigo.

Obtidas a estrutura da rede e as probabilidades condicionadas, a rede bayesiana pode ser empregada para realizar interpretações e diagnósticos usando um ou mais dados novos. Para tanto é necessário avaliar a influência destes dados em medidas de probabilidades associadas aos diferentes nós. Este é o objetivo dos algoritmos de propagação e fusão de influências, comentados a seguir.

3. Fusão de Influências e Propagação de Mensagens numa Rede Probabilística

A partir da atribuição de valores específicos a um conjunto de variáveis correspondentes aos dados de entrada disponíveis, processo denominado de instanciação, deseja-se calcular seu impacto nas probabilidades "*a posteriori*" do conjunto de variáveis (nós) que compõem uma rede bayesiana previamente estabelecida.

A distribuição de probabilidade a posteriori associado a cada nó ao longo do processamento será aqui denominada "distribuição induzida" (em Inglês é usado o termo *belief*).

Ao instanciar um nó da rede (atribuir valor à variável que esse nó representa), essa informação é enviada aos nós vizinhos que irão iniciar um processo de propagação multi-dimensional até o equilíbrio ser alcançado.

O algoritmo proposto por Pearl [5] é o mais conhecido para avaliar a fusão de influências. Esse algoritmo trabalha com a suposição que a rede não possua laços. Um laço em uma rede bayesiana consiste em um caminho fechado no grafo obtido a partir da estrutura desta rede após a substituição dos seus arcos por arcos não direcionados. Por exemplo, a rede apresentada na Figura 1 consiste de um laço e o nó C é dito ser o **nó extremo** desse laço.

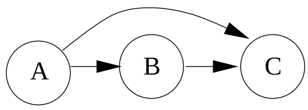

Figura 1: Laço numa rede bayesiana.

A existência de laços deve ser evitada pois, devido ao mecanismo de propagação de mensagens, é possível que uma informação que circule em um laço chegue a um mesmo nó por dois caminhos diferentes. Isto pode levar a erros na avaliação das influências dos dados e também produzir instabilidade nesta avaliação.

Para redes que apresentam laços, outros algoritmos têm sido propostos para avaliar a fusão de influências [6], [7]. Essas técnicas se caracterizam essencialmente por fazer alterações na estrutura da rede, seja através da fusão de nós ou do desdobramento de um nó em dois ou mais, objetivando no primeiro caso retirar os ciclos e no segundo para fazer a propagação mesmo com ciclos.

Outros algoritmos usam a remoção de arcos da rede para "quebrar" os laços [8]. As técnicas propostas para a retirada de arcos foram inicialmente propostas para simplificar a rede e assim diminuir sua complexidade. Entretanto, se o arco eliminado pertencer a um laço, a sua remoção implica naturalmente a retirada deste laço.

O procedimento proposto em [8] consiste em, partindo da rede original e da distribuição conjunta obtida a partir das distribuições condicionadas, remover um dos arcos dessa rede, avaliar a nova distribuição conjunta e em seguida calcular a divergência entre esta e a distribuições conjunta original. Os arcos cujas retiradas proporcionam menores divergências são candidatos a serem removidos. Neste trabalho são ainda apresentados critérios adicionais para escolha dos arcos que serão removidos.

Uma das desvantagens do procedimento anterior é que, para cada arco candidato à remoção, a distribuição de probabilidades da nova rede sem esse arco deve ser obtida.

A seguir, é apresentado um novo critério para retirada de arcos, mais voltado para a remoção de laços, que usa a estrutura da rede antes da obtenção da distribuição de probabilidades final.

4. Critério para a Retirada de Laços em Redes Bayesianas

O método de remoção de arcos proposto em [8] requer necessariamente várias construções de distribuições de probabilidades.

Admitindo-se o uso de um método bayesiano para a etapa de construção da rede, a estrutura de rede obtida da base de dados é a que minimiza a métrica utilizada (no caso do algoritmo K3, a métrica de descrição). Parece portanto razoável utilizar como critério para escolha de uma rede sem arcos que se aplique à mesma base de dados, o grau de "proximidade"entre esta e a rede ótima, utilizando-se a mesma métrica empregada pelo algoritmo de construção da rede.

Esse raciocínio pode ser usado para obter um algoritmo de eliminação de laços que dispense as construções de novas distribuições de probabilidades e assim produza uma significativa redução de complexidade em relação ao procedimento proposto em [8]. Com base nesta idéia e na métrica utilizada pelo algoritmo K3, propõe-se a seguir um novo algoritmo para eliminação de laços.

O objetivo do algoritmo é buscar a rede sem laços derivada da rede originalmente construída com o algoritmo K3, através da retirada de arcos cuja métrica DL mais se aproxima do valor ótmo desta métrica, para a base de dados em questão.

Aplicando localmente este raciocínio a cada laço a ser quebrado, a escolha de um arco para ser removido com o objetivo de eliminar este laço deveria recair sobre o arco candidato cuja retirada produzisse a menor alteração na métrica DL da rede. Propõe-se a seguir que esta remoção de arcos seja orientada por um critário de entropia condicionada.

A entropia condicionada $H(X \mid Y)$ entre duas variáveis aleatórias discretas X e Y é dada pela Equação 1 abaixo e pode ser vista como uma medida de incerteza sobre a primeira variável quando se tem algum conhecimento da segunda [3]. Assim, quanto maior $H(X \mid Y)$ maior a incerteza sobre X dado que se conhece Y.

$$H(X \mid Y) = -\sum_{\mathcal{X}} \sum_{\mathcal{Y}} P(X,Y) \log(P(X \mid Y)) \quad (1)$$

Dados dois arcos que chegam em um nó Na provenientes dos nós Nb e Nc, propõe-se usar para escolha do arco a ser eliminado o critério de maior entropia condicionada. Se $H(Na \mid Nb) > H(Na \mid Nc)$, o conhecimento de Nb mantém a incerteza sobre Na em um nível maior do que quando se tem conhecimento de Nc. Portanto, se for necessário eliminar um dos arcos que chegam a Na propõe-se que seja eliminado o que liga os nós Na e Nb. As distribuições de probabilidades necessárias para o cálculo das entropias condicionadas podem ser obtidas dos dados.

Cabe notar que a aplicação do critério de escolha acima apresentado se reflete diretamente no valor da métrica DL da rede. Note-se que na construção da rede original o algoritmo K3 só acrescenta um arco entre dois nós se o mesmo diminuir o valor da métrica DL, o que implica que o referido arco representa uma ligação que traz muita da informação entre essas variáveis. De forma similar, quando da aplicação do critério acima proposto para eliminação de laços, retirar o arco de maior entropia condicionada entre os nós por ele ligados implica remover a ligação que traz menor informação entre os nós envolvidos.

A identificação de laços presentes na rede pode ser feita usando o algoritmo *depth-first search*. Note-se que a remoção de um arco pode quebrar mais de um laço. Para se obter um procedimento eficiente (e sub-ótimo) de busca dos laços a serem quebrados, propõe-se o tratamento sequencial e isolado de cada nó, no sentido de identificar e eliminar laços. Desta forma, após a retirada dos arcos necessários para quebrar todos os laços formados por um dado nó, o nó seguinte é testado quanto à existência de nós, e assim sucessivamente, até que o penúltimo nó seja testado.

O procedimento completo para remoção de arco que aqui se propõe é sintetizado a seguir na forma de um algoritmo, onde N_{nos} é a quantidade de nós da rede.

<u>Algoritmo para retirada de laços:</u>

1. Para $i = 1...N_{nos} - 1$

2. Aplique o algoritmo *depth-first search* para verificar se existe um laço.

2.1 Se existir, vá ao nó extremo desse laço e encontre as entropias condicionadas entre esse nó e os seus antecessores que pertencem ao laço. Retire o arco que liga os nós de maior entropia condicionada e volte para 2.

2.2 Aplique o algoritmo *depth-first* ao nó i para determinar se ainda existe laços

2.2.1 Caso não existam mais laços que o nó i pertença, i = i+1 e vá para 2

2.2.2 Caso contrário volte para 2.1

A seguir este procedimento para remoção de laços é aplicado à rede obtida no projeto [1] que deu origem ao presente trabalho.

5. Aplicação à Rede Miastenia Grave

Para a obtenção da rede a ser usada na análise das doenças neuromusculares foram selecionados sete sintomas. Esses sintomas serão os nós da rede, denominada rede Miastenia Grave. Para tanto foi montada uma base de dados com 46 casos. Esses casos foram em parte fornecidos pelo Prof. Jovany Medeiros, da Universidade Estadual da Paraíba, e em parte obtidos por buscas na Internet e na literatura especializada.

A partir da base de dados foi aplicado o algoritmo K3 descrito anteriormente. A rede obtida está descrita na Tabela 1, onde são mostrados os nós e os seus pais. Na Figura 2 tem-se uma representação parcial dessa rede (não são mostradas todos os ramos), onde pode ser observada a ocorrência de laços.

Como exemplo da operação do algoritmo proposto e de sua relação com o critério MDL, considere-se a Figura 3, que exibe o laço formado pelos nós $N1$, $N2$ e $N3$. Na Tabela 2 estão apresentados os valores das entropias condicionadas $H(Ni \mid Nj)$ entre os nós Ni e Nj e os valores das métricas DL obtidas quando se retira o

Nós	Pais
N1	Não tem pais
N2	N1
N3	N1 e N2
N4	N1, N2 e N3
N5	N1, N3 e N4
N6	N1, N3, N4 e N5
N7	N1, N3, N4, N5 e N6

Tabela 1: Nós da rede bayesiana para a Miastenia Gravis e seus respectivos conjuntos de pais

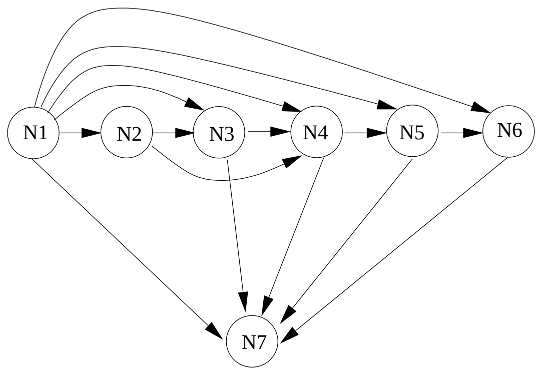

Figura 2: Representação parcial da rede da Tabela 1.

arco entre Ni e Nj. Observa-se que a retirada do arco entre os nós $N3$ e $N1$ leva a métrica DL inicial de 222,90 para 232,32 enquanto a retirada do arco que liga $N2$ e $N3$ leva esta métrica para o valor 224,56. Assim sendo, é melhor retirar o segundo arco do que o primeiro, sob o aspecto do menor aumento da métrica DL. A retirada deste arco implica a quebra do laço em questão.

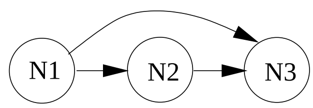

Figura 3: Exemplo de Nós da rede Miastenia Gravis que formam um laço.

Aplicando o algoritmo proposto para a remoção de arcos à rede Miastenia Gravis, obtém-se a rede derivada apresentada na Figura 4. O teste desta rede como instrumento efetivo para realização de diagnósticos será objeto de trabalho futuro.

6. Conclusões

Propôs-se um algoritmo simples para eliminação de laços em redes bayesianas através da retirada seletiva de arcos, o qual se baseia na entropia condicionada entre os nós conectados pelos arcos candidatos à remoção. A relação entre este critério e o critério MDL utilizado para construção de redes bayesianas associadas a bases de dados foi discutida de forma qualitativa, destacando-se a consonância entre os mesmos. Foi apresentado um ex-

		DL
$H(N_3 \mid N_1)$	0,763	232,32
$H(N_3 \mid N_2)$	0,933	224,56

Tabela 2: Entropias condicionadas e valores da métrica DL após a retirada dos arcos.

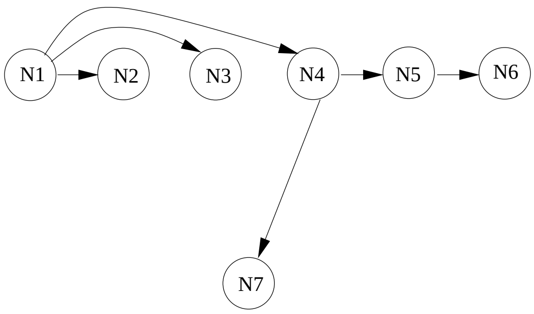

Figura 4: Rede obtida após a quebra dos laços da rede apresentada na Figura 2.

emplo de aplicação deste algoritmo a uma base de dados relativos à doença Miastenia Grave. A eficácia de utilização da rede resultante para diagnóstico desta doença será avaliada em trabalho posterior. Serão também consideradas alternativas para se lidar com a possibilidade ocorrência de mais de uma eliminação de laços pela retirada de um mesmo nó.

Referências

[1] Francisco M. de Assis, Ernesto L. Pinto e Jovany L. A. de Medeiros. *Sistema Computacional para Auxílio ao Diagnóstico de Doenças Neuromusculares*. Universidade Federal de Campina Grande, Campina Grande, Brasil, 2001.

[2] Remco R. Bouckaert. *Probabilistic Network Construction Using the Minimum Description Length Principle*. Department of Computer Science, Utrecht University, Utrecht, The Netherlands, 1994.

[3] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, USA, 1991.

[4] Edwards Herskovits. *Computer-Based Probabilistic-Network Construction*. Univesity of California, California, USA, 1991.

[5] Judea Pearl. “Fusion, Propagation, and Structuring in Belief Networks”. *Artificial Intelligence*, 29:241–288, May 1986.

[6] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publisher, San Francisco, USA, 1988.

[7] F. J. Díez. Local conditioning in Bayesian networks. *Artificial Intelligence*, 87:1–20, 1996.

[8] Robert A. Engelen. “Approximating Bayesian Belief Network by Arc Removal”. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 19(8):916–920, August 1997.