

UNIVERSIDADE DO EXTREMO SUL CATARINENSE – UNESC

CURSO DE CIÊNCIA DA COMPUTAÇÃO

EDROALDO LUMMERTZ DA ROCHA

**O FORMALISMO BAYESIANO PELO ALGORITMO DE PASSAGEM DE MENSAGENS DE
JUDEA PEARL NA SHELL PEGASUS UNCERTAINTY MODELING**

CRICIÚMA, JULHO DE 2008

EDROALDO LUMMERTZ DA ROCHA

**O FORMALISMO BAYESIANO PELO ALGORITMO DE PASSAGEM DE MENSAGENS DE
JUDEA PEARL NA SHELL PEGASUS UNCERTAINTY MODELING**

Trabalho de Conclusão de Curso apresentado
para obtenção do Grau de Bacharel em Ciência da
Computação da Universidade do Extremo Sul
Catarinense.

Orientadora: Prof^a.MSc. Priscyla Waleska
Targino de Azevedo Simões

CRICIÚMA, JULHO DE 2008

Aos meus pais, Maria e Eriovaldo, pelo seu amor, exemplo, dedicação e incentivo. Ao meu irmão Adriano, por seu carisma. Amo muito vocês.

AGRADECIMENTOS

Agradeço a Jesus Cristo, por ser um revolucionário autêntico, Alguém que amou e ama incondicionalmente, capaz de inspirar corações, mudar vidas e prover esperança a um mundo perdido e sedento por amor. Muito obrigado por guiar meus passos e me consolar nos momentos difíceis, me ensinar que o conhecimento desassociado da sabedoria não significa nada e por me ajudar a chegar ao fim de mais uma jornada, mas principalmente, fornecer os direcionamentos para o futuro.

Com muito amor, agradeço aos meus pais e ao meu irmão, por serem meus amores. Quero agradecer minha mãe, por ser um exemplo de liderança, persistência, criatividade, alegria, sabedoria e principalmente amor. Mãe, eu quero lhe agradecer por ser uma mulher de fibra, corajosa, inspirada e por transferir seu desejo de mudar o mundo para o meu coração. Obrigado por ser um exemplo de vitórias e caráter, obrigado por existir, palavras não são suficientes para descrevê-la. Te amo muito mãe.

Agradeço ao meu pai por sua ousadia e visão. Por ser criativo e por me transferir a curiosidade de saber como e porque as coisas funcionam. Com o tempo, aprendi a dar valor ao seu exemplo. Obrigado por ser um exemplo de que mudar é possível. Te amo muito pai.

Agradeço ao meu irmão, pelo seu carisma, alegria, bondade e por ser a pessoa com quem aprendi e me espelhei no aspecto de convívio social. Te amo mano. Agradeço a vocês três por fazerem parte de minha vida, por me ensinarem tantas coisas e por contribuir para tornar-me a pessoa que sou hoje.

Agradeço a minha namorada, Angel, por ser um verdadeiro anjo em minha vida. Por todo seu amor, compreensão, paciência, doçura e carinho. Por me ajudar a amenizar a saudade da minha família, por ser linda e inspiradora. Muito obrigado por

estar comigo nesta etapa e em etapas futuras. Te amo muito Angel.

Agradeço ao REAJA¹, especialmente ao Paulo Santos, Simone Campos (Pérola), e o Alex, por ser um movimento revolucionário no qual meu inconformismo pode ser direcionado e aplicado a mudança efetiva do mundo em que vivemos. Muito obrigado pelo seu amor e dinamismo.

Agradeço aos amigos que fiz em Criciúma e por toda a convivência que tivemos oportunidade de desfrutar.

Agradeço a minha professora e orientadora, pelos conselhos e principalmente por sua amizade. Também agradeço por ela proporcionar a oportunidade de contribuir com o grupo de pesquisa em uma área de meu interesse. Muito obrigado por acreditar em mim.

Enfim, muito obrigado a todos que contribuíram para minha formação de vida e profissional. Também agradeço ao Grupo de Profissionais Diferenciados da UNESC. Um abraço a todos.

Meus agradecimentos ao Grupo de Pesquisa em Informática Médica e Telemedicina – Projeto Kiron, e ao Grupo de Pesquisa em Inteligência Computacional Aplicada. Agradeço também ao Programa Institucional de Bolsas de Iniciação Científica (PIBIC) por financiar esta pesquisa.

¹ <http://www.reaja.com.br>

“Nas ciências naturais, a Natureza nos dá um mundo, e temos apenas que descobrir as suas leis. Em computação, somos nós que introduzimos as leis e criamos um mundo.”

(Alan Kay)

RESUMO

A mudança paradigmática ocorrida na matemática e ciência do século XXI proporcionou o desenvolvimento de diversas teorias para modelagem da incerteza, como fatores de certeza, lógica *fuzzy*, teoria de Dempster-Shafer e redes bayesianas. A partir destas, diversas ferramentas para automatização da resolução de problemas que apresentam o conceito de incerteza foram desenvolvidas. No entanto, a evolução destes diferentes ambientes de desenvolvimento ocorreu de forma não integrada, culminando em ambientes específicos e que consideram um único formalismo para modelagem da incerteza. Portanto, objetivando fornecer as quatro teorias supracitadas, uma ferramenta intitulada *shell Pegasus Uncertainty Modeling* começou a ser desenvolvida pelo Grupo de Pesquisa em Inteligência Computacional Aplicada, sendo a atribuição desta pesquisa a modelagem da *shell Pegasus* e o desenvolvimento da teoria de redes bayesianas nesta ferramenta. A construção do módulo bayesiano da *shell Pegasus* possibilitou construir e verificar como um modelo gráfico caracteriza um mecanismo eficiente para representação de conhecimento e como as variáveis no domínio influenciam umas as outras explorando relacionamentos de dependência e independência condicional em grafos, mecanismo capaz de otimizar o processo de inferência, a qual baseia-se em redes com estrutura de poliárvore, devido a uma representação compacta da distribuição de probabilidades da rede bayesiana.

Palavras-Chaves: Modelagem da Incerteza, Probabilidade, Redes Bayesianas, Inferência, Poliárvores, Algoritmo Passagem de Mensagens de Pearl.

ABSTRACT

The paradigm shift occurred in mathematics and science of the XXI century provided the development of various theories for uncertainty modeling, as certainty factors, fuzzy logic, theory of Dempster-Shafer and Bayesian networks. From these, a variety of tools to automate the resolution of problems that present the concept of uncertainty were developed. However, the evolution of these different development environments occurred on a non-integrated form, culminating in specific environments and that consider a single formalism for uncertainty modeling. Therefore, to provide the four theories above, a tool called shell Pegasus Uncertainty Modeling began to be developed by the Research Group in Applied Computational Intelligence and the objectives of this research is accomplish the modeling of shell Pegasus and the development of the theory of Bayesian networks in this tool. The construction of the Bayesian module of the shell Pegasus make possible to build and verify how a graphic model features an efficient mechanism for representation of knowledge and how the variables in the domain influence each other exploring relationships of dependence and independence conditional on graphs, mechanism capable of optimizing the process the inference, which is based on networks structured as polytrees due to a compact representation of the distribution of probabilities of Bayesian network.

Keywords: Uncertainty Modeling, Probability, Bayesian Networks, Inference, Polytrees, Pearl Message Passing Algorithm.

LISTA DE ILUSTRAÇÕES

Figura 1. Arquitetura de um sistema fuzzy	25
Figura 2. Uma rede bayesiana.....	42
Figura 3. O nó D divide a poliárvore em duas poliárvores desconexas.....	47
Figura 4. Uma poliárvore	51
Figura 5. Variáveis envolvidas na atualização de λ	54
Figura 6. Interface gráfica do Analytica	60
Figura 7. Interface gráfica da <i>shell</i> Bayesian.....	61
Figura 8. Interface gráfica do GeNie.....	62
Figura 9. Interface gráfica da ferramenta Hugin Expert	64
Figura 10. Interface gráfica da <i>shell</i> Netica	66
Figura 11. Método <i>initial_net()</i>	69
Figura 12. Método <i>update_tree()</i>	70
Figura 13. Método <i>send_λ_msg()</i>	71
Figura 14. Método <i>send_π_msg</i>	72
Figura 15. Rede usada no exemplo	73
Figura 16. Rede Bayesiana Inicializada.....	77
Figura 17. Probabilidades da rede bayesiana após a evidência g1.....	80
Figura 18. Diagrama de atividades da <i>shell</i> Pegasus	82
Figura 19. Diagrama de casos de uso da <i>shell</i> Pegasus.....	83
Figura 20. Diagrama de classes do módulo bayesiano	84
Figura 21. Diagrama de casos de uso da ferramenta de redes bayesianas	85
Figura 22. Diagrama de atividades da ferramenta de redes bayesianas.....	86
Figura 23. Interface inicial da ferramenta.....	91

Figura 24. Barra de ferramentas para trabalhar com redes bayesianas	91
Figura 25. Barra para inicializar os atributos do nó	92
Figura 26. Rede bayesiana para o diagnóstico de lombalgia.	92
Figura 27. Rede bayesiana inicializada.....	93
Figura 28. Probabilidades quando nenhuma evidência é fornecida.....	94
Figura 29. Modelo bayesiano desenvolvido com o nó S0 selecionado	94
Figura 30. Probabilidades quando histórico familiar positivo	95
Figura 31. Rede bayesiana quando histórico familiar positivo	95
Figura 32. Probabilidades quando histórico familiar positivo e dor aguda.....	95
Figura 33. Modelo quando histórico familiar positivo e dor aguda.....	96
Figura 34. Rede usada no exemplo	116
Figura 35. Método <i>initial_tree()</i>	117
Figura 36. Método <i>send_pi_msg()</i>	118
Figura 37. Método <i>send_pi_msg()</i>	119
Figura 38. Método <i>update_tree()</i>	120
FFigura 39. Método <i>send_lambda_msg()</i>	121
Figura 40. Método <i>send_lambda_msg()</i>	122

LISTA DE TABELAS

Tabela 1. Variáveis aleatórias de uma população com 500 pessoas	34
Tabela 2. Probabilidades associadas a uma população de 500 pessoas	35
Tabela 3. Variáveis, seus estados e respectivos significados.....	43
Tabela 4. Variáveis, estados e seus significados.....	73
Tabela 5. Variáveis relacionadas ao paciente	88
Tabela 6. Variáveis dispostas segundo os hábitos de vida.....	89
Tabela 7. Variáveis organizadas pelo critério dor.....	89
Tabela 8. Variáveis organizadas por exame físico.....	89
Tabela 9. Variáveis agrupadas segundo o exame neurológico	90
Tabela 10. Variáveis, estados e seus significados.....	116

LISTA DE ABREVIATURAS E SIGLAS

CPT	Condition Probability Table
TPC	Tabela de Probabilidade Condicional
UNESC	Universidade do Extremo Sul Catarinense
UML	Unified Modeling Language
API	Application Programming Interface

SUMÁRIO

1 INTRODUÇÃO.....	16
1.1 OBJETIVO GERAL	17
1.2 OBJETIVOS ESPECÍFICOS	18
1.3 JUSTIFICATIVA.....	18
1.4 ESTRUTURA DO TRABALHO.....	21
2 MODELAGEM DA INCERTEZA	23
2.1 LÓGICA FUZZY	24
2.2 FATORES DE CERTEZA	26
2.3 TEORIA DE DEMPSTER-SHAFER	27
2.4 REDES BAYESIANAS	28
3 FUNDAMENTOS DE PROBABILIDADE.....	30
3.1 FUNÇÕES E ESPAÇOS DE PROBABILIDADE.....	30
3.2 AXIOMAS DA PROBABILIDADE.....	31
3.3 VARIÁVEIS ALEATÓRIAS	32
3.4 DISTRIBUIÇÕES DE PROBABILIDADE CONJUNTA, MARGINAL E CONDICIONAL	34
3.5 TEOREMA DA PROBABILIDADE TOTAL.....	36
3.6 MODELO BAYESIANO DE PROBABILIDADE.....	37
3.7 TEOREMA DE BAYES.....	38
3.8 INDEPENDÊNCIA	38
3.9 INDEPENDÊNCIA CONDICIONAL	39

4 REDES BAYESIANAS	40
4.1 POR QUE UTILIZAR REDES BAYESIANAS?	44
5 INFERÊNCIA EM REDES BAYESIANAS	46
5.1 INFERÊNCIA EM POLIÁRVORES.....	47
5.1.1 Algoritmo de passagem de mensagens de Pearl.....	48
5.1.1.1 Atualizando λ	53
5.1.1.2 Atualizando π	56
5.1.1.3 Resumo do algoritmo de propagação em poliárvores	57
6 TRABALHOS CORRELATOS	59
6.1 ANALYTICA	59
6.2 BAYESIANLAB	61
6.3 GRAPHICAL NETWORK INTERFACE (GENIE)	62
6.4 HUGIN EXPERT	63
6.5 NETICA	65
7 O FORMALISMO BAYESIANO NA SHELL PEGASUS UNCERTAINTY MODELING.....	67
7.1 ESTUDO DAS <i>SHELLS</i> DISPONÍVEIS	68
7.2 DEMONSTRAÇÃO MATEMÁTICA DO MECANISMO DE INFERÊNCIA.....	68
7.3 MODELAGEM DA SHELL	81
7.4 IMPLEMENTAÇÃO DO MÓDULO BAYESIANO E DA SHELL PEGASUS	87
7.5 APLICAÇÃO DO MÓDULO BAYESIANO A UMA PROBLEMÁTICA DA ÁREA DA SAÚDE .	87
7.6 RESULTADOS OBTIDOS	90

8 CONCLUSÃO	98
REFERÊNCIAS	100
ANEXO A - DIAGNÓSTICO DA LOMBALGIA	103
8.1 DOENÇAS EXISTENTES E DOENÇAS DIAGNOSTICADAS	103
8.2 SINAIS E SINTOMAS ENVOLVIDOS NO DIAGNÓSTICO	105
ANEXO A - ARTIGO	107
APÊNDICE A – MODELAGEM MATEMÁTICA DO ALGORITMO DE INFERÊNCIA.....	116

1 INTRODUÇÃO

As técnicas, métodos e conceitos de Inteligência Artificial têm sido cada vez mais utilizados na resolução de problemas complexos (FERRARI, 2005), nos quais a aplicação da computação algorítmica convencional é inviável, como reconhecimento de padrões, análise de séries temporais, diagnóstico médico, entre outros (BRAGA, 2000).

Dentre as técnicas de inteligência artificial frequentemente utilizadas na resolução de problemas destacam-se os sistemas especialistas, que devido sua habilidade de manipular conhecimento, são capazes de reproduzir o comportamento de especialistas humanos em determinado domínio de aplicação (BITTENCOURT, 2001).

Devido a sua reconhecida aplicabilidade, estes sistemas tornaram-se objetos de estudo no âmbito do desenvolvimento de ferramentas que possibilitam sua construção de forma facilitada. Tais ferramentas intituladas *shells*, implementam conceitos de inteligência artificial, fornecendo uma abstração das características fundamentais dos sistemas especialistas, por meio da separação entre a base de conhecimento, da máquina de inferência e dos demais módulos que compõe sua arquitetura, possibilitando a reutilização desta em diferentes domínios de aplicação. Além disso, provê metodologias para representação de conhecimento e modelagem da incerteza, como por exemplo, redes *bayesianas*, lógica *fuzzy*, entre outras.

No entanto, há uma carência de *shells* que implementam duas ou mais técnicas de inteligência artificial para modelar a incerteza em suas diversas manifestações, como imprecisão, aleatoriedade, não especificidade, entre outras.

Neste sentido, o Grupo de Pesquisa em Inteligência Computacional Aplicada teve a iniciativa de desenvolver uma *shell*, que envolva diferentes métodos para modelagem da incerteza, denominada Pegasus *Uncertainty Modeling*, que serão

disponibilizados inicialmente, no que se refere à inferência, em quatro módulos: Fatores de Certeza, Teoria de Dempster-Shafer, Lógica *Fuzzy* e Redes Bayesianas.

As redes bayesianas podem ser utilizadas como metodologia para representação de conhecimento segundo um enfoque probabilístico constituindo base de conhecimento dos sistemas especialistas probabilísticos, sendo seu mecanismo de inferência baseado no Teorema de Bayes. Destaca-se ainda que tais sistemas são freqüentemente utilizados na área da saúde no auxílio à tomada de decisão e diagnóstico médico, porém não se restringindo a mesma, podendo também ser aplicadas nas demais áreas de conhecimento existentes.

Neste contexto, esta pesquisa propõe o desenvolvimento de uma *shell*, por meio da inferência bayesiana, para construção automatizada de sistemas especialistas probabilísticos, que será integrada a *shell* Pegasus, pois esta abordagem possibilita a modelagem de relacionamentos probabilísticos utilizando redes bayesianas (NEOPOLITAN, 2004), em domínios de aplicação onde a incerteza por aleatoriedade se faz presente.

1.1 OBJETIVO GERAL

Desenvolver inferência bayesiana na *shell* Pegasus *Uncertainty Modeling*

1.2 OBJETIVOS ESPECÍFICOS

- a) Compreender a modelagem da incerteza por aleatoriedade;
- b) Entender as redes bayesianas;
- c) Aplicar as redes bayesianas para modelagem de relacionamentos probabilísticos;
- d) Utilizar método de inferência bayesiana a ser definido;
- e) Demonstrar matematicamente o funcionamento do método de inferência bayesiano utilizado;
- f) Proporcionar uma ferramenta para modelagem de sistemas especialistas probabilísticos;
- g) Disponibilizar uma API.

1.3 JUSTIFICATIVA

Nos últimos anos, as pesquisas em desenvolvimento de ferramentas computacionais para inteligência artificial, que visam capturar e simular o comportamento de especialistas humanos, intensificaram-se (FERRARI, 2005), dando origem a uma ampla variedade de sistemas que visam atender a diversos domínios de aplicação.

Nestes domínios, freqüentemente encontra-se o conceito de incerteza, o qual se destaca como uma das diversas mudanças paradigmáticas ocorridas na ciência e matemática deste século (KLIR, 1995). Na ciência, esta mudança foi manifestada por uma transição gradual do ponto de vista tradicional, que insiste que a incerteza é

indesejável e deveria ser evitada, para uma visão alternativa, que considera a incerteza útil e aplicável. Ao modelar sistemas, de um modo geral, permitir mais incerteza tende a reduzir a complexidade e aumentar a credibilidade do modelo resultante (KLIR, 1995).

Conseqüentemente, há a necessidade de modelá-la eficientemente. Para tanto, foram desenvolvidos modelos matemáticos cujo objetivo é representar a incerteza em suas possíveis e diversas manifestações. A Teoria dos Fatores de Certeza, a Teoria de Dempster-Shafer, a Lógica *Fuzzy* e as Redes Bayesianas, são exemplos de esforços em direção a sua modelagem eficaz.

Atualmente existem muitas ferramentas que implementam estas teorias oferecendo um mecanismo automatizado para construir sistemas inteligentes aplicados a diferentes áreas do conhecimento humano onde a modelagem da incerteza é essencial, como medicina, robótica, entre outras.

No entanto, estas ferramentas, livres ou proprietárias, implementam somente um esquema de representação de conhecimento e conseqüentemente, restringem sua área de aplicação. Dependendo do problema, é difícil definir precisamente qual o tipo de incerteza que se está modelando. Um mesmo problema pode ser abordado segundo um enfoque probabilístico, *fuzzy* ou ainda utilizando abordagens híbridas.

Portanto, uma ferramenta que possibilite a escolha do modelo de representação de conhecimento juntamente com diferentes mecanismos de inferência seria de grande valia para a comunidade de inteligência artificial, munindo-a de uma ferramenta que promova a automatização do desenvolvimento de sistemas inteligentes onde a modelagem da incerteza é essencial.

Assim, esta pesquisa compreende o desenvolvimento de uma *shell* para construção de sistemas especialistas probabilísticos, que será integrada a *shell* Pegasus, pois os modelos de inferência bayesianos, devido sua capacidade de modelar

relacionamentos probabilísticos mostram-se adequados para resolução de problemas em diversas áreas com a presença de incerteza por aleatoriedade.

As redes bayesianas são esquemas de representação de conhecimento, baseadas na teoria da probabilidade, utilizadas para construir a base de conhecimento de sistemas especialistas probabilísticos (KORB; NICHOLSON, 2004). Utilizando o conceito de probabilidade condicional, elas sugerem um mecanismo unidirecional para modelagem da causalidade tornando-as ideais para expressar relacionamentos causais, ou seja, baseados na relação causa-efeito. Este tipo de raciocínio é bastante utilizado em diversas áreas, principalmente, no processo de diagnóstico médico, uma vez que especialistas no domínio raciocinam probabilisticamente baseados nos sintomas apresentados por um paciente.

Sob este prisma, a *shell* resultante do desenvolvimento desta pesquisa pode vir a auxiliar diversas áreas de atuação do conhecimento humano, podendo ser, por exemplo, utilizada para auxiliar no desenvolvimento de uma aplicação para diagnóstico de doenças cardiovasculares ou para detecção de objetos em aplicações de realidade aumentada.

Em um contexto mais amplo, o módulo bayesiano integrado a *shell* Pegasus, pode possibilitar o desenvolvimento e análise de metodologias híbridas para a modelagem da incerteza, onde o raciocínio bayesiano coexista com o raciocínio *fuzzy*, por exemplo, ou outro formalismo, considerando-se que a utilização destes métodos pode produzir resultados tão satisfatórios quanto sua utilização separadamente.

Neste sentido, busca-se com esta pesquisa o desenvolvimento de uma ferramenta que possibilite automatizar o processo de desenvolvimento de sistemas especialistas probabilísticos.

1.4 ESTRUTURA DO TRABALHO

Esta pesquisa é constituída por 7 capítulos. No capítulo 1 são apresentadas as motivações para seu desenvolvimento bem como a definição do problema, justificativa e objetivos.

O capítulo 2 disserta sobre a modelagem da incerteza, suas origens e teorias atuais para tratá-la. É fornecida uma breve descrição sobre lógica *fuzzy*, fatores de certeza, teoria de Dempster-Shafer e redes bayesianas, explanação esta justificada pelo desenvolvimento de uma ferramenta que visa as respectivas implementações, sendo o assunto desta pesquisa, a teoria de redes bayesianas sob o enfoque da criação de um mecanismo de inferência nestas estruturas.

Neste raciocínio, o capítulo 3 apresenta os fundamentos de probabilidade necessários para compreensão da teoria de redes bayesianas e do mecanismo de inferência desenvolvido, descritos nos capítulos 4 e 5, respectivamente.

O capítulo 4 aborda as redes bayesianas, sua definição, em que formalismo matemático são baseadas, como sua estrutura torna o processo de inferência computacionalmente eficiente e motivos pelos quais utilizá-las.

O algoritmo de inferência desenvolvido é apresentado no capítulo 5, cuja implementação é baseada na troca de mensagens entre os diferentes nós que constituem o modelo para propagação das distribuições de probabilidades quando da ocorrência de uma evidência. O formalismo matemático é detalhadamente descrito e as equações utilizadas para distribuir as probabilidades no modelo são derivadas, demonstrando a origem das mesmas e contribuindo para uma compreensão aprofundada do mecanismo de inferência.

No capítulo 6 são apresentadas algumas ferramentas disponíveis para redes bayesianas e o capítulo 7 descreve o trabalho desenvolvido, que consiste na modelagem, tanto da *shell* Pegasus quanto do módulo de redes bayesianas, bem como sua implementação. Também descreve a utilização da ferramenta desenvolvida para prover uma aplicação de suporte à decisão no diagnóstico de lombalgia, além da modelagem matemática do algoritmo de inferência, a partir da qual é possível visualizar, por meio de um exemplo, a aplicação do pseudocódigo descrito neste capítulo e conseqüentemente, o funcionamento da inferência bayesiana utilizando um mecanismo que explora a estrutura de uma poliárvore para propagar evidências.

Finalizando, o capítulo 8 apresenta as conclusões e recomendações de trabalhos futuros.

2 MODELAGEM DA INCERTEZA

De um modo geral, a resolução de problemas é abordada em termos de sistemas construídos a partir de modelos. Ao projetar sistemas, a maximização de sua utilidade pode ser obtida por meio da ponderação entre três características fundamentais de cada modelo de sistema: complexidade, credibilidade e incerteza (COSTA, 1999). Embora este relacionamento ainda não seja completamente compreendido, reconhece-se a relevância da incerteza para aumentar sua utilidade e, portanto, permitir mais incerteza tende a reduzir a complexidade e aumentar a credibilidade do modelo resultante, tornando-o possivelmente mais manipulável e útil (KLIR, 1995; COSTA, 1999).

A incerteza é então um conceito útil para construção de sistemas que pode ser manipulada para possibilitar o desenvolvimento de modelos maximamente úteis com relação ao propósito para os quais são construídos (COSTA, 1999). Portanto, procura-se desenvolver métodos por meio dos quais o nível de incerteza estimado seja tal que proporcione um aumento da credibilidade e uma redução da complexidade do modelo de sistema construído (KLIR, 1995).

Este conceito foi responsável por uma mudança paradigmática na ciência e matemática deste século cuja manifestação resultou em um processo de transição do ponto de vista tradicional, que considera a incerteza indesejável e que a mesma deveria ser evitada, para o moderno que a considera uma entidade relevante e aplicável em domínios complexos, baseada em dois estágios (KLIR, 1995).

O primeiro estágio teve início no século XIX com a introdução de métodos estatísticos e probabilísticos no estudo de processos em nível molecular, anteriormente realizados por meio da mecânica newtoniana². Estes métodos então ultrapassaram os

² Teoria relacionada ao movimento dos corpos publicada por Isaac Newton em 1687 no livro

limites dos processos moleculares e foram estendidos para distintas ramificações da ciência contribuindo para diminuir a distância existente entre diversas áreas do pensamento humano (COSTA, 1999).

O segundo estágio tornou-se explícito na literatura em 1960, sendo caracterizado pelo surgimento de diversas teorias cujo objetivo é fornecer mecanismos formais para o tratamento de diferentes tipos de incerteza (KLIR, 1995), como imprecisão e ignorância. Dentre as teorias emergentes neste período destaca-se a Lógica *Fuzzy*, Fatores de Certeza, Dempster-Shafer e as Redes Bayesianas.

2.1 LÓGICA FUZZY

A teoria dos conjuntos *fuzzy* é uma extensão da teoria clássica dos conjuntos que possibilita expressar transições graduais da pertinência a não-pertinência de determinado elemento a determinado conjunto (WANG, 1997). Nesta teoria um mesmo elemento pode pertencer simultaneamente a diferentes conjuntos, com diferentes graus de pertinência, possibilitando uma modelagem realística de diversos problemas frequentemente encontrados na área de inteligência artificial.

Esta capacidade fornece não somente um mecanismo de representação de incerteza significativo e poderoso, como também um formalismo para representar matematicamente conceitos qualitativamente vagos ou imprecisos expressos em linguagem natural, que possui vasta aplicação em diversas áreas do conhecimento humano (KLIR, 1995).

A partir desta teoria são desenvolvidos sistemas *fuzzy*, os quais são

freqüentemente aplicados nas áreas de controle, tomada de decisão, robótica, visão computacional, entre outras. Estes sistemas são constituídos por uma arquitetura modular, conforme ilustra a Figura 1, passível de modificação em função do propósito para o qual o sistema é projetado (REZENDE, 2005).

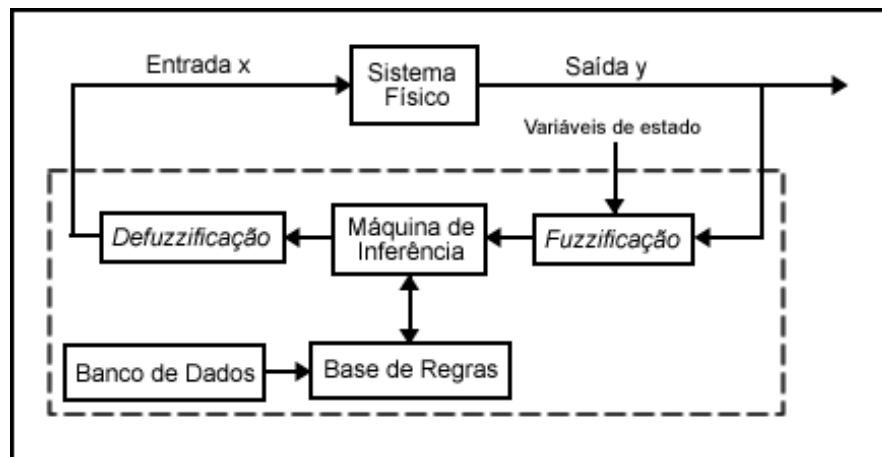


Figura 1. Arquitetura de um sistema fuzzy

Fonte: Adaptado de ALMEIDA, P; EVSUKOFF, A. (2005).

A fuzzificação refere-se a um processo de transformação onde entradas *crisp*³ são transformadas em entradas *fuzzy*. Primeiramente são determinadas as funções de pertinência para cada entrada do sistema, seguindo-se à definição das funções, o processo recebe valores de entrada que são comparados com as funções de pertinência armazenadas, dando origem as entradas *fuzzy* (OLIVEIRA JUNIOR; CALDEIRA, 2007).

A função de pertinência é um critério que define o quanto um elemento pertence a determinado conjunto. Este valor define o grau para o qual um elemento é similar ou compatível com o conceito representado por um conjunto *fuzzy*, ou seja, indivíduos podem pertencer em maior ou menor grau a determinado conjunto conforme expresso por seu grau de pertinência (BADIRU; CHEUNG, 2002).

A base de regras *fuzzy* representa o conhecimento necessário sobre o

³ Para distinguir um conjunto *fuzzy* de um conjunto clássico (não *fuzzy*), o último passa a ser denominado conjunto *crisp*, e os elementos pertencentes a este conjunto, entradas *crisp* (KLIR, 1995).

problema, sendo constituída por duas partes fundamentais: antecedente e conseqüente. A medida de adequação de cada regra é definida de acordo com o conectivo utilizado para implementar a premissa das regras (COX, 2005). Já o termo conseqüente utiliza-se de tal medida sendo que seus valores são integrados por intermédio do uso da **T-norma** ou de uma **T-conorma**⁴ adequadas.

A defuzzificação é o procedimento pelo qual um conjunto de saída *fuzzy* é transformado em um valor *crisp*.

2.2 FATORES DE CERTEZA

Os fatores de certeza são uma representação comum para pesos **heurísticos** onde números maiores que 0 são utilizados para evidências positivas e números menores que 0 são usados para evidências negativas. Estes números são apenas heurísticas e nenhum critério de correteza teórica está associado a eles (BADIRU; CHEUNG, 2002).

Essa teoria foi inicialmente utilizada no sistema especialista MYCIN (DAVIS; KING, 1984), desenvolvido para manipular incertezas, diagnosticar e recomendar terapias para infecções no sangue. A cada regra do sistema é associado um fator de certeza, um número entre -1 e 1, que indica o grau de certeza na ocorrência desta regra.

Um fator de certeza é dividido em duas partes, a Medida de Crença (MC) e a Medida de Descrença (MD). A MC representa a confiança em uma afirmação tendo ocorrido uma determinada evidência, e a MD mostra o grau de descrença na mesma

⁴ A lógica *fuzzy* utiliza os operadores generalizados AND e OR da lógica clássica. Eles são chamados T-Normas (para AND) e T-Conormas (para OR) (KLIR, 1995).

afirmação com a ocorrência da evidência (COSTA; SIMÕES, 2004).

2.3 TEORIA DE DEMPSTER-SHAFER

A teoria de Dempster-Shafer visa a distinção entre ignorância e incerteza, uma vez que as mesmas são essencialmente diferentes (BADIRU; CHEUNG, 2002). Ao invés de calcular a probabilidade de uma proposição, calcula a probabilidade de que a evidência admita a proposição.

Esta medida de crença é chamada função de crença e é representada por $Bel(x)$ (RUSSEL; NORVIG, 2004). Não conhecer um valor específico de uma variável não necessariamente implica que a variável é assunto do campo da incerteza. Com a teoria da probabilidade clássica, é necessário considerar crença e descrença como funções opostas, ou seja, se A, B, C são os únicos três eventos contidos no espaço amostral e considerando $P(A)=0.3$, $P(B) = 0.6$, a probabilidade estimada do evento C, para satisfazer o teorema da probabilidade total, seria igual a $P(C)=0.1$.

No entanto, esta pode não ser uma representação adequada para o raciocínio humano e a probabilidade de C pode não ser objeto de estudo da teoria da probabilidade ou da área de modelagem da incerteza, mas sim a ignorância no grau de ocorrência de C (BADIRU; CHEUNG, 2002). Desta forma, o conhecimento das probabilidades A e B não implica na possibilidade de inferir C.

Nesta teoria, o grau de crença em uma proposição é dado pelo intervalo de dois valores, sendo que um mede o grau de crença e o outro o grau de plausibilidade. Esta abordagem rejeita o princípio de que a ausência a favor de uma proposição significa a atribuição de evidência à sua negação (COSTA; SIMÕES, 2004).

Por outro lado, a teoria da evidência trabalha com conjuntos de proposições, o que permite lidar de forma explícita com a incerteza por ignorância.

Esta teoria parte de um conjunto de proposições Q , denominado universo finito de proposições. Por exemplo, em um contexto de diagnóstico médico, são atribuídas probabilidades aos subconjuntos de proposições e não às proposições (COSTA; SIMÕES, 2004).

2.4 REDES BAYESIANAS

Uma Rede Bayesiana é um formalismo baseado na teoria dos grafos e na teoria da probabilidade que possibilita a representação gráfica do conhecimento incerto e a propagação de probabilidades em sua estrutura por meio de algoritmos de inferência (CASTILHO; GUTIÉRREZ; HADI,1998). Elas possuem vasta aplicação em diversas áreas do conhecimento humano, principalmente em contextos médicos (MARQUES; DULTRA, 2007).

Elas oferecem uma estrutura eficiente para propagação de probabilidades explorando as relações de independência condicional existentes entre os nós do grafo possibilitando uma representação fatorizada da distribuição de probabilidade conjunta e conseqüentemente tornando o processo de inferência em modelos probabilísticos computacionalmente tratável (LUNA, 2004).

No decorrer deste capítulo, foram abordadas diferentes teorias para raciocínio sob incerteza, as quais enquadram-se em contextos específicos, podendo tanto ser utilizadas isoladamente ou em arquiteturas híbridas, dependendo do problema envolvido.

Especificamente, este trabalho utiliza redes bayesianas como uma

metodologia para resolução de problemas que apresentam um tipo distinto de incerteza cuja modelagem pode ser realizada baseada em teoria dos grafos e princípios de probabilidade. Portanto, para a compreensão da teoria de redes bayesianas é necessário, primeiramente, uma introdução aos principais conceitos de probabilidade envolvidos.

3 FUNDAMENTOS DE PROBABILIDADE

Ao iniciar o estudo de redes bayesianas é necessário, inicialmente, introduzir alguns conceitos relacionados à teoria das probabilidades para compreender o funcionamento dos algoritmos de inferência e aprendizado.

O objetivo deste capítulo é fornecer uma introdução sobre os fundamentos da probabilidade, como as funções e os espaços de probabilidade (seção 3.1), seus axiomas (seção 3.2), a definição de variáveis aleatórias (seção 3.3), distribuições de probabilidade (seção 3.4), os teoremas da probabilidade total marginal e condicional (seção 3.5), o modelo bayesiano de probabilidade (seção 3.6), o teorema de Bayes (seção 3.7) e os relacionamentos de independência (seção 3.8) e independência condicional (seção 3.9).

3.1 FUNÇÕES E ESPAÇOS DE PROBABILIDADE

A teoria da probabilidade trabalha com experimentos que apresentam um conjunto distinto de resultados (NEOPOLITAN, 2004). Exemplos destes experimentos incluem escolher a carta do topo de um baralho com 52 cartas, com os 52 resultados correspondentes a 52 cartas diferentes, ou ainda escolher uma pessoa em uma determinada população e determinar se a pessoa faz graduação em ciência da computação com os dois resultados possíveis sendo “sim” ou “não”.

Matematicamente, o conjunto de resultados possíveis de um experimento é denominado espaço amostral. Em um espaço amostral finito, cada subconjunto do mesmo é denominado evento. Um subconjunto contendo exatamente um evento é denominado evento elementar. Uma vez identificado o espaço amostral, uma função de

probabilidade pode ser definida como (NEOPOLITAN, 2004):

Definição 1.1 Suponha um espaço amostral S contendo n elementos distintos, ou seja, $S = \{e_1, e_2, \dots, e_n\}$.

Uma função que atribui um número real $P(E)$ a cada evento $E \subseteq U$ é chamada função de probabilidade do conjunto se satisfaz as seguintes condições (NEOPOLITAN, 2004):

- a) $0 \leq P(\{e_i\}) \leq 1$ para $1 \leq i \leq n$
- b) $P(\{e_1\}) + P(\{e_2\}) + \dots + P(\{e_n\}) = 1$
- c) para cada evento $E = \{e_1, e_2, \dots, e_n\}$ que não é um evento elementar

$$P(E) = P(\{e_{i1}\}) + P(\{e_{i2}\}) + \dots + P(\{e_{ik}\}).$$

O par (S, P) é chamado espaço de probabilidade (NEOPOLITAN, 2004). Este conceito de probabilidade pode ser intuitivamente visualizado conforme ilustra o exemplo: Considerando o experimento de obter uma carta do topo de um baralho com 52 cartas, então U contém as faces das 52 cartas. Utilizando o princípio da indiferença, a probabilidade da ocorrência de cada evento $e \in U$ é $P(\{e\}) = 1/52$.

O princípio da indiferença diz que eventos elementares são considerados equiprováveis se não há uma razão para preferir um ou outro. De acordo com este princípio, quando há n eventos elementares, a probabilidade de cada um deles é caracterizada pelo quociente $1/n$. No entanto, esta não é a única noção de probabilidade (NEOPOLITAN, 2004).

3.2 AXIOMAS DA PROBABILIDADE

Dado um experimento aleatório qualquer descrito pelo espaço amostral S , a

teoria da probabilidade é baseada em uma função $P: S \rightarrow [0,1]$ que mapeia os subconjuntos de S em um número real pertencente ao intervalo $[0,1]$ representada por $P(A)$, denominada probabilidade do evento A , que satisfaz os seguintes axiomas (BARBETTA; REIS; BORNIA, 2004):

a) $P(S) = 1$

b) $0 \leq P(E) \leq 1$

c) se E_1, E_2, \dots, E_n são eventos mutuamente exclusivos, então:

$$P(E_1 \cup E_2 \cup \dots \cup E_n) = P(E_1) + P(E_2) + \dots + P(E_n)$$

Muitas propriedades podem ser obtidas a partir dos axiomas acima (da SILVA; LADEIRA, 2002):

a) $P(\emptyset) = 0$

b) se $E \subseteq F \subseteq S$ então $P(E) \leq P(F)$

c) se $E, F \subseteq S$, então $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

3.3 VARIÁVEIS ALEATÓRIAS

Dado um espaço de probabilidade (S, P) , uma variável aleatória é uma função em S , isto é, ela apresenta uma característica funcional $X: S \rightarrow S_X$ que associa cada elemento do espaço amostral S a um elemento de seu contradomínio S_X (LUNA, 2004).

Por exemplo, seja E o experimento de lançar duas moedas e verificar o número de caras (H); X é uma variável aleatória que indica o número de caras obtidas no experimento E ; $S_X = \{0,1,2\}$ o conjunto de valores possíveis de X ; e $S = \{HH, HT, TH, TT\}$ o espaço amostral S de E . Considerando-se então a natureza funcional do conceito de variável aleatória $X: S \rightarrow S_X$, obtém-se $X(HH) = 2$, $X(HT) = X(TH) = 1$ e $X(TT) = 0$.

Este aspecto funcional é comumente ignorado e considera-se que uma variável aleatória X assume um dos valores possíveis de seu contradomínio, que passa a ser denominado domínio de X e a probabilidade passa a ser definida neste domínio (SILVA; LADEIRA, 2002), por exemplo, $P(X = 0) = P(TT) = 1/4$, $P(X = 1) = P(HT \text{ ou } TH) = 2/4$ e $P(X = 2) = P(HH) = 1/4$. De um modo geral, se E for qualquer evento de S , então $P(E) = \sum_i P(S_i)$.

Estatisticamente, uma variável aleatória X pode ser definida como um símbolo representando um conjunto de valores, denominado espaço de X , que por simplicidade, pressupõem-se ser contável. Estas variáveis representam alguma característica da entidade modelada e não se conhece *a priori* o valor associado a esta característica. Para resolver esta incerteza, são desenvolvidos relacionamentos probabilísticos entre as variáveis, assumindo que tais relacionamentos são os mesmos encontrados no modelo, caso contrário, o mesmo não é aplicável (NEOPOLITAN, 2004).

Em um contexto de diagnóstico de câncer, por exemplo, poderia haver uma variável aleatória, denominada câncer, cujo espaço é {ausente, presente}. Considerar câncer = presente implica em afirmar que a entidade sendo modelada possui câncer, no entanto, matematicamente, esta expressão é uma primitiva utilizada em expressões de probabilidade (NEOPOLITAN, 2004).

Variáveis aleatórias são comumente representadas por letras maiúsculas acompanhadas de um índice subscrito. Por exemplo, as variáveis idade e sexo podem ser representadas por X_1 e X_2 , respectivamente. Os valores associados às variáveis aleatórias são representados por letras minúsculas, por exemplo, x_1^j , x_1^a , x_1^t , para uma variável idade, onde os índices sobrescritos denotam jovem, adulto e terceira idade, respectivamente e para a variável sexo, x_2^m e x_2^f , designam sexo masculino e feminino

respectivamente (DIEZ, 2003).

3.4 DISTRIBUIÇÕES DE PROBABILIDADE CONJUNTA, MARGINAL E CONDICIONAL

Probabilidade conjunta: Seja um conjunto de n variáveis aleatórias $V = \{X_1, X_2, \dots, X_n\}$ de modo que X_i possui um espaço infinitamente contável. Uma função que atribui um número real $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ a cada combinação de valores dos x_i 's tal que o valor de x_i é escolhido a partir do espaço de X_i é chamada de distribuição de probabilidade conjunta das variáveis aleatórias em V se satisfaz as seguintes condições (NEOPOLITAN, 2004):

a) para cada combinação de valores de x_i 's,

$$0 \leq P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \leq 1$$

b) é satisfeita a condição

$$\sum_{x_1, x_2, \dots, x_n} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = 1$$

Probabilidade marginal: Dado uma distribuição de probabilidade conjunta $P(x_1, x_2, \dots, x_n)$, a distribuição marginal para uma variável X_i é obtida pelo somatório das probabilidades para todas as configurações possíveis para as demais variáveis (DIEZ, 2003).

$$P(x_i) = \sum_{x_j | X_j \neq X_i} P(x_1, x_2, \dots, x_n)$$

Considerando, por exemplo, a Tabela 1, que apresenta dados relacionados a uma população de 500 pessoas, tendo como variáveis aleatórias o sexo: masculino e feminino, e a idade, tem-se:

Tabela 1. Variáveis aleatórias de uma população com 500 pessoas

Intervalo de Idade	Masculino	Feminino	Total
<18	67	68	135
18-65	122	126	248
>65	57	60	117
Total	246	254	500

Fonte: DIEZ, F. (2003)

Realizando então um experimento sobre esta população, que consiste em escolher aleatoriamente um indivíduo e considerando que o mesmo satisfaça o princípio da indiferença, a probabilidade que uma pessoa possua certa idade e determinado sexo é obtida pela expressão: (número de pessoas com a mesma faixa etária e número de pessoas com determinado sexo) / número de indivíduos da amostra. A Tabela 2 ilustra a distribuição de probabilidade conjunta e marginal das variáveis envolvidas neste exemplo.

Tabela 2. Probabilidades associadas a uma população de 500 pessoas

P	Masculino	Feminino	Total
<18	$P(x_1^j, x_2^m) = 0,134$	$P(x_1^j, x_2^f) = 0,136$	$P(x_1^j) = 0,270$
18-65	$P(x_1^a, x_2^m) = 0,244$	$P(x_1^a, x_2^f) = 0,252$	$P(x_1^a) = 0,496$
>65	$P(x_1^t, x_2^m) = 0,114$	$P(x_1^t, x_2^f) = 0,120$	$P(x_1^t) = 0,234$
Total	$P(x_2^m) = 0,492$	$P(x_2^f) = 0,508$	1,000

Fonte: DIEZ, F. (2003)

As probabilidades marginais são obtidas somando-se as linhas para X_1 e as colunas para X_2 . Esta probabilidade também poderia ser obtida utilizando a Tabela 1, obtendo o número total da primeira linha, 135 e dividindo pelo tamanho da população, ou seja, $135/500 = 0,270$ e conseqüentemente, a soma das probabilidades dos valores

de cada variável é 1.

Uma vez definidos os conceitos de probabilidade conjunta e marginal, define-se como probabilidade condicional, denotada por $P(X|Y)$, a probabilidade da ocorrência do evento X condicionada ao conhecimento do evento Y, sendo descrita pela Equação 1 (BARRETO, 2001):

$$P(X|Y) = \frac{P(x, y)}{P(y)} \quad (1)$$

Continuando o exemplo anterior, a probabilidade de uma pessoa do sexo masculino possuir idade superior a 65 anos é a probabilidade de sua idade ser maior que 65 (x_1^t) dado que o indivíduo é do sexo masculino (x_2^m): $P(x_1^t | x_2^m) = P(x_1^t, x_2^m) / P(x_2^m) = 0,114 / 0,492 = 0,23171$.

A equação 1 pode ser reescrita como $P(x, y) = P(X|Y)P(y)$ obtendo a chamada regra do produto, a qual pode ser generalizada para obter a regra da cadeia. Dado que $X = \{X_1, X_2, \dots, X_n\}$ é um conjunto de variáveis aleatórias, a distribuição de probabilidade conjunta de X é dada pela equação (KORB; NICHOLSON, 2004):

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | x_1, \dots, x_{i-1})$$

3.5 TEOREMA DA PROBABILIDADE TOTAL

Supondo n eventos E_1, E_2, \dots, E_n tal que $E_i \cap E_j = \emptyset$ e $E_1 \cup E_2 \cup \dots \cup E_n = S$. Estes eventos são denominados mutuamente exclusivos e exaustivos. Logo, o teorema da probabilidade total diz que para qualquer outro evento F (NEOPOLITAN, 2004),

$$P(F) = \sum_{i=1}^n P(F \cap E_i)$$

Se $P(E_i) \neq 0$, então $P(F \cap E_i) = P(F|E_i)P(E_i)$. Logo, se $P(E_i) \neq 0$ para todo i , o teorema da probabilidade total é freqüentemente aplicado na forma (PEARL, 1988):

$$P(F) = \sum_{i=1}^n P(F|E_i)P(E_i)$$

3.6 MODELO BAYESIANO DE PROBABILIDADE

Um estudante de ciência da computação está prestes a realizar uma prova de inteligência artificial, especificadamente sobre redes bayesianas. Considerando que ele estudou e tem conhecimento sobre o conteúdo, ele sabe que ao estimar suas chances de tirar uma boa nota em 98%, ele está utilizando um conceito subjetivo ou bayesiano de probabilidade, ou seja, sua crença é de 98%.

Uma vez que ele tenha estudado, ele sabe que esta probabilidade não é um quociente $1/n$ e também não é baseada em freqüências relativas, ou mesmo em uma estimativa de freqüência relativa, uma vez que o evento de realizar a prova não pode ser repetidamente executado nas mesmas condições. Esta probabilidade representa então a crença de tirar uma boa nota na prova de inteligência artificial, sendo denominada grau de crença ou probabilidade subjetiva.

Observa-se com este exemplo um relacionamento causal expresso pela probabilidade condicional $P(\text{Nota=Boa}|\text{Estudou=Sim}) = 98\%$, ou seja, a probabilidade é condicionada pelo conhecimento prévio do evento Estudou e pode ser calculada utilizando o teorema de Bayes.

3.7 TEOREMA DE BAYES

O teorema de Bayes é o mecanismo básico segundo o qual a inferência probabilística é realizada. Sua derivação advém do teorema da multiplicação e sua equação é descrita pela Equação 2 (RUSSELL; NORVIG, 2004):

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)} \quad (2)$$

Prova: Utilizando a definição de probabilidade condicional:

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \quad \text{e} \quad P(F|E) = \frac{P(F \cap E)}{P(E)}$$

Multiplicam-se as igualdades pelo denominador do lado direito e devido a ambos os termos $P(E \cap F)$ e $P(F \cap E)$ serem iguais, obtém-se.

$$P(E|F)P(F) = P(F|E)P(E)$$

Divide-se então a equação acima por $P(F)$ e obtém-se o teorema de Bayes.

3.8 INDEPENDÊNCIA

Dois eventos E e F são independentes se satisfazem as seguintes condições (NEOPOLITAN, 2004):

- a) $P(E|F) = P(E)$ e $P(E) \neq 0, P(F) \neq 0$
- b) $P(E) = 0$ ou $P(F) = 0$

Esta definição indica que dois eventos são independentes mesmo que sejam baseados na probabilidade condicional de E dado F, devido à independência ser simétrica. Ou seja, se $P(E) \neq 0$ e $P(F) \neq 0$, então $P(E|F) = P(E)$, se e somente se $P(F|E) = P(F)$. Se dois eventos E e F são independentes então $P(E,F) = P(E)P(F)$

(NEOPOLITAN, 2004).

Exemplo: Sendo $E = \{reiCopas, reiEspadas, rainhaCopas\}$, $F = \{reiCopas, reiPaus, rainhaCopas\}$, $G = \{reiCopas, reiEspadas, reiPaus, reiOuros\}$, então:

$$P(E) = 3/52$$

$$P(E|F) = 2/3$$

$$P(E|G) = 1/2$$

$$P(E|F, G) = 1/2$$

Logo, E e F não são independentes, mas são condicionalmente independentes dado G. Assim, independência condicional é um conceito importante para a inferência bayesiana, pois reduz a complexidade do algoritmo.

3.9 INDEPENDÊNCIA CONDICIONAL

Dois eventos E e F são condicionalmente independentes dado G se $P(G) \neq 0$ se satisfazem as seguintes condições (NEOPOLITAN, 2004):

a) $P(E|F, G) = P(E|G)$ e $P(E|G) \neq 0$, $P(F|G) \neq 0$

b) $P(E|G) = 0$ ou $P(F|G) = 0$

c) Se dois eventos E e F são condicionalmente independentes dado G então

$$P(E, F|G) = P(E|G)P(F|G)$$

Este capítulo descreveu alguns conceitos necessários para a compreensão das redes bayesianas. Foram abordados o conceito de probabilidade e suas distribuições, variáveis aleatórias, axiomas, relacionamentos de independência, entre outros. A informação contida neste capítulo será útil para a compreensão do algoritmo de inferência descrito nos próximos capítulos e fornecerá subsídios para modelagem probabilística a partir de redes bayesianas.

4 REDES BAYESIANAS

Redes bayesianas são grafos acíclicos dirigidos constituídos por dois componentes fundamentais: a) um componente estrutural, S , que define um relacionamento qualitativo causal entre os nós de um grafo, e b) parâmetros numéricos, θ , que quantificam a relação probabilística causal existente entre os nós de S (LUNA, 2004) (WILLIANSO, 2005).

Considerando que em um relacionamento causal a causa precede o efeito, há a necessidade de um mecanismo unidirecional para modelagem da causalidade, sendo o conceito de probabilidade condicional indicativo desses relacionamentos sem conexão direta (NASSAR, 2003).

Neste sentido, as redes bayesianas fornecem um modelo matemático que expressa causalidade explorando as relações de independência entre as variáveis no domínio sendo modelado e possibilitando a fatorização da distribuição de probabilidade conjunta. Esta distribuição permite responder a diversas questões sobre um domínio de dados. No entanto, sua dificuldade de representação aumenta em função do número de variáveis.

Na aplicação do teorema de Bayes, a exploração de independências condicionais entre variáveis aleatórias possibilita simplificar os cálculos e reduzir o número de probabilidades condicionais que precisam ser especificadas. Sob este aspecto, as redes bayesianas, utilizando o teorema de Bayes como mecanismo de inferência, representam as dependências entre as variáveis aleatórias existentes bem como fornecem uma representação compacta da distribuição de probabilidade conjunta, caracterizando uma estrutura eficiente para a realização do procedimento de inferência (PEARL, 1988).

Formalmente, uma rede bayesiana representa uma distribuição de probabilidade conjunta P sobre um conjunto de variáveis aleatórias X . Assim, uma rede bayesiana em X é constituída por uma estrutura S , que codifica assertivas de independência condicional sobre as variáveis X e um conjunto, instância de θ , de distribuições locais associadas às variáveis em X . Juntas, a estrutura S e o conjunto de parâmetros θ definem uma distribuição de probabilidade conjunta tal que (SILVA; LADEIRA, 2002) (LUNA, 2004):

- a) S é um grafo acíclico dirigido;
- b) Os nós em S estão em uma relação de 1-1 (um nó em S , corresponde a uma variável em X) com as variáveis em X ;
- c) Cada variável X_i , em X , denota uma variável e também o respectivo nó em S ;
- d) Pa_i denota os nós pais de X_i e também as variáveis correspondentes a esses pais.

$$P(x) = \prod_{i=1}^n P(x_i | pa_i) \quad (3)$$

- e) A distribuição conjunta de X é dada pela equação 3, onde x é uma instância de X e pa_i é uma instância de Pa_i .

Para ilustrar alguns dos conceitos abordados, considera-se um exemplo adaptado de Neopolitan (2004) relacionado à modelagem probabilística de um contexto médico. Em uma aplicação que considera que uma variável influencia diretamente outra variável da mesma entidade, um possível exemplo seria a presença ou ausência de uma doença influenciando diretamente o resultado do teste. Por décadas o teorema de Bayes foi utilizado para realizar inferências probabilísticas nestas situações, por meio do cálculo de probabilidade condicional (NEOPOLITAN, 2004).

No entanto, supondo um contexto mais complexo, onde diferentes variáveis

estão relacionadas, um indivíduo, por exemplo, com histórico de fumante apresenta uma influência direta na ausência ou presença de bronquite e se possui ou não um câncer de pulmão. Então, a presença ou ausência de cada uma destas doenças apresenta uma influência direta no fato do indivíduo apresentar ou não fadiga, e a presença de câncer de pulmão indica uma influência direta se o exame de raio-X do tórax é positivo ou não, cujos relacionamentos são ilustrados na Figura 2.

Neste contexto o objetivo é realizar inferência probabilística utilizando variáveis que não estão diretamente relacionadas. Por exemplo, calcular a probabilidade condicional de bronquite e câncer de pulmão dado que o indivíduo fuma, está fatigado e apresentou raio-X do tórax positivo. As redes bayesianas resolvem este problema caracterizando uma estrutura gráfica para representação de distribuições de probabilidades condicionais (JENSEN, 2001).

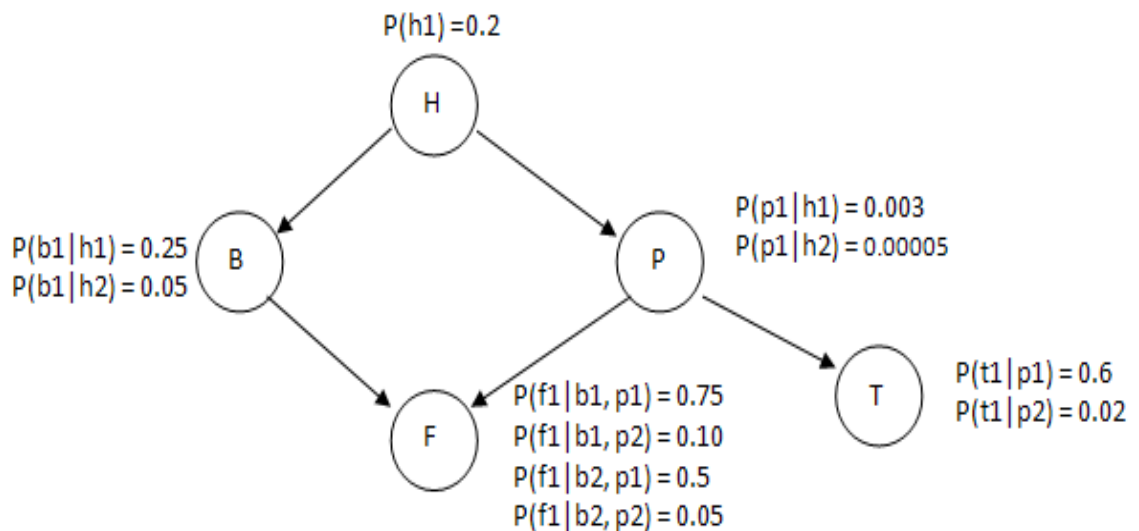


Figura 2. Uma rede bayesiana
 Fonte: NEOPOLITAN, R. (2004)

Tabela 3. Variáveis, seus estados e respectivos significados

Variável	Valor	Descrição
H	h1	Há histórico de fumante
	h2	Não há histórico de fumante
B	b1	Tem bronquite
	b2	Não tem bronquite
P	p1	Tem câncer de pulmão
	p2	Não tem câncer de pulmão
F	f1	Tem fadiga
	f2	Não tem fadiga
T	t1	Raio-X do tórax positivo
	t2	Raio-X do tórax negativo

Fonte: NEOPOLITAN, R. (2004)

Conforme ilustra a Figura 2, os arcos representam influências diretas, por exemplo, a existência de um arco de H para P indica que um histórico de fumante tem influência direta na presença de câncer de pulmão. Similarmente, há um arco conectando P a T pois a presença de câncer tem influência direta no resultado de raio-X do tórax. No entanto, não há um arco conectando diretamente H a T porque o histórico de fumante somente influencia o resultado do raio-X (T) por meio de P, ou seja, não há um relacionamento causal direto entre as variáveis H e T.

Uma rede bayesiana pode ser construída de diferentes formas. Uma delas se dá pela criação dos arcos que representam influências diretas entre as variáveis. Uma vez que a rede tenha sido qualitativamente especificada, a parte quantitativa é então definida utilizando probabilidades condicionais dos valores de cada variável, dado cada combinação de valores das variáveis dos pais na rede, exceto dos nós raízes, que

possuem probabilidades *a priori*. No momento em que o modelo foi tanto qualitativamente quanto quantitativamente especificado, a inferência bayesiana pode ser realizada, possibilitando, por exemplo, conhecer os valores das probabilidades condicionais de bronquite e câncer de pulmão dado que é constatado um histórico de fumante e um teste de raio-X positivo (NEOPOLITAN, 2004).

4.1 POR QUE UTILIZAR REDES BAYESIANAS?

Redes bayesianas possibilitam expressar assertivas de independência em uma estrutura gráfica intuitiva. Além disso, representam e armazenam distribuições de probabilidades conjuntas de uma forma compacta explorando relações de independência condicional entre as variáveis e, devido à redução do número de probabilidades condicionais que precisam ser especificadas (LUNA, 2004), devido à estrutura S satisfazer a propriedade de Markov⁵ (KORB; NICHOLSON, 2004), o procedimento de inferência torna-se computacionalmente eficiente.

Além disso, dependendo do tipo de aplicação, existem características que tornam as redes bayesianas preferíveis em relação a outros modelos (LUNA, 2004):

- a) aquisição de conhecimentos: possibilidade de utilizar, em um mesmo modelo, conhecimentos de naturezas essencialmente distintas, como probabilidades subjetivas, estatísticas ou equações;
- b) representação de conhecimentos: o conhecimento é explicitamente expresso em sua estrutura gráfica, tornando sua interpretação direta, eficiente e intuitiva. Esta característica facilita a validação do modelo e

⁵ Esta propriedade diz que não há dependências diretas no sistema sendo modelado que não sejam explicitamente mostradas por meio de arcos (KORB; NICHOLSON, 2004).

sua utilização;

- c) utilização de conhecimentos: um mesmo modelo pode ser utilizado para efetuar diferentes tipos de raciocínio, como preditivo, diagnóstico ou ambos.
- d) ferramentas disponíveis: existe uma variedade de ferramentas que possibilitam a construção de redes bayesianas, abstraindo a complexidade da teoria adjacente a estes modelos e possibilitando seu desenvolvimento utilizando componentes gráficos intuitivos, além de permitir a execução de mecanismos de inferência e/ou aprendizado.

Uma vez compreendidos os conceitos relacionados às redes bayesianas bem como o entendimento de como sua estrutura gráfica possibilita representar de forma objetiva assertivas de dependência e independência condicional, o próximo capítulo aborda o processo de inferência realizado nestes modelos, por meio do algoritmo de passagem de mensagens apresentado por Judea Pearl 1988.

5 INFERÊNCIA EM REDES BAYESIANAS

A inferência bayesiana possibilita estimar a probabilidade das variáveis aleatórias pertencentes a um domínio de aplicação modelado a partir da ocorrência de um conjunto de evidências E (CASTILHO; GUTIÉRREZ; HADI, 1998). Sua importância está relacionada não somente a obtenção de distribuições de probabilidade condicionais como também para aprendizagem automática de redes bayesianas.

Em contextos médicos, por exemplo, o diagnóstico para determinada doença é obtido analisando-se os sintomas (evidências) apresentados pelo paciente. Este mecanismo é denominado propagação de evidências (inferência) e consiste em atualizar as probabilidades das variáveis quando novas informações são fornecidas.

Existem diferentes algoritmos de inferência, os quais podem ser classificados em três categorias fundamentais (NEOPOLITAN, 2004): algoritmos exatos, aproximados e simbólicos. Os algoritmos exatos calculam as probabilidades das variáveis sem outro erro senão aquele produzido pelas limitações de representação numérica do processador (CASTILHO; GUTIÉRREZ; HADI, 1998).

No entanto, em situações onde os métodos exatos não são aplicáveis, utilizam-se algoritmos aproximados, os quais aplicam diferentes técnicas de simulação para obter probabilidades aproximadas. Os algoritmos simbólicos podem ser utilizados tanto com parâmetros numéricos quanto simbólicos, obtendo probabilidades na forma simbólica em função dos parâmetros (LUNA, 2004).

Assim, este capítulo descreve o algoritmo de inferência bayesiana apresentado por Judea Pearl em 1988 (PEARL, 1988) para poliárvores, que são grafos nos quais existe no máximo, um caminho não dirigido entre dois nós (NEOPOLITAN, 2004). São descritos a inferência em redes bayesianas com estrutura de poliárvores e

derivadas as equações para realização deste processo, bem como um resumo das fórmulas utilizadas no algoritmo.

5.1 INFERÊNCIA EM POLIÁRVORES

Uma rede bayesiana pode ser representada por um modelo gráfico denominado poliárvore. Em uma poliárvore, dois nós estão conectados por apenas um caminho, fazendo com que cada nó subdivida esta estrutura em duas poliárvores desconexas, uma contendo os pais e outra os filhos do respectivo nó, além dos subgrafos associados a cada pai e filho do nó em questão (CASTILHO; GUTIÉRREZ; HADI, 1998).

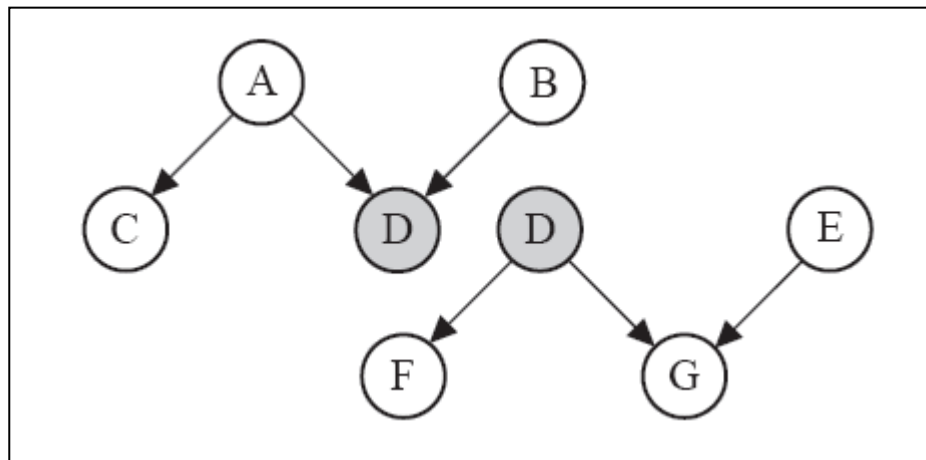


Figura 3. O nó D divide a poliárvore em duas poliárvores desconexas
 Fonte: CASTILHO, E.; GUTIÉRREZ, J.; HADI, A. (1998).

Por exemplo, na Figura 3, o nó D divide a poliárvore em duas poliárvores não conexas. A primeira, constituída pelos nós {A, B, C}, inclui seus pais e os nós acessíveis desde D por meio destes. A segunda, formada pelos nós {E, F, G}, inclui seus filhos e os nós acessíveis desde D por meio de seus filhos. Conforme ilustrado, o

nó D separa estes dois conjuntos e, portanto, a relação de independência condicional $I(\{A, B, C\}\{E, F, G\}|D)$ é graficamente constatada.

Neste tipo de grafo, a inferência bayesiana pode ser eficientemente realizada utilizando passagens de mensagens⁶ entre os distintos subgrafos da poliárvore, por meio de computações locais⁷ (NEOPOLITAN, 2004). Portanto, o algoritmo de passagens de mensagens (PEARL, 1988) pode ser utilizado para propagar as probabilidades pela rede quando um conjunto de evidências E é fornecido.

5.1.1 Algoritmo de passagem de mensagens de Pearl

Supondo que se conheça a evidência $E = e$, o objetivo da inferência probabilística é calcular as probabilidades condicionais $P(x|e)$ para todos os valores x de um nó X que não estejam contidos em E . Para facilitar o cálculo destas probabilidades, o conjunto de evidências é decomposto em dois subconjuntos disjuntos, cada um dos quais contendo as poliárvores separadas por X na poliárvore original. Portanto, E pode ser subdividido em (CASTILHO; GUTIÉRREZ; HADI, 1998):

- a) e_X^+ , que é o subconjunto de E acessível desde X por meio de seus pais.
- b) e_X^- , que é o subconjunto de E acessível desde X por meio de seus filhos.

Portanto, $e_X^+ \cup e_X^- = E$ e $e_X^+ \cap e_X^- = \emptyset$. Uma vez definidos estes conceitos, inicia-se a derivação das equações para realização de inferência probabilística utilizando poliárvores (PEARL, 1988) (NEOPOLITAN, 2004).

$$P(x|e) = P(x|e_X^-, e_X^+)$$

Utilizando o teorema de Bayes, Equação 2:

⁶ Mecanismo pelo qual as probabilidades são distribuídas pela rede bayesiana (NEOPOLITAN, 2004).

⁷ Cada nó que compõe a rede faz cálculos localmente, baseados em sua informação e na informação proveniente de outros nós, que uma vez calculada é transmitida para outros nós, por meio de uma mensagem (KORB; NICHOLSON, 2004).

$$P(x|e) = \frac{P(e_X^-, e_X^+ | x) \cdot p(x)}{P(e_X^-, e_X^+)}$$

Segundo Diez (2003) utilizando a propriedade para variáveis condicionalmente independentes $P(e_X^-, e_X^+ | x) = P(e_X^- | x)P(e_X^+ | x)$ obtém-se:

$$P(x|e) = \frac{P(e_X^- | x)P(e_X^+ | x) \cdot p(x)}{P(e_X^-, e_X^+)}$$

Aplicando o teorema de Bayes ao termo $P(e_X^+ | x)$ da equação anterior e eliminado os termos comuns ao numerador e ao denominador da equação, obtém-se a seqüência de cálculos abaixo:

$$P(x|e) = \frac{P(e_X^- | x)P(x|e_X^+) \cdot p(e_X^+) \cdot P(x)}{P(x)P(e_X^-, e_X^+)}$$

$$P(x|e) = \frac{P(e_X^- | x)P(x|e_X^+) \cdot p(e_X^+) \cdot \cancel{P(x)}}{\cancel{P(x)}P(e_X^-, e_X^+)}$$

$$P(x|e) = \alpha P(e_X^- | x)P(x|e_X^+)$$

Onde:

$$\alpha = [P(e_X^- | e_X^+)]^{-1}$$

é uma constante de normalização obtida utilizando a definição de probabilidade condicional, ou seja, ao eliminar os termos da equação acima descrita, quando as variáveis são isoladas, obtém-se $P(e_X^- | x)P(x|e_X^+)$ e $p(e_X^+) / P(e_X^-, e_X^+)$, portanto, aplicando a Equação (1) nesta última fórmula e tomando seu inverso, a constante α pode ser obtida.

$$P(x|e) = \alpha \lambda(x) \pi(x) \tag{4}$$

A Equação (4), portanto, é utilizada para atualizar as probabilidades de x quando evidências são fornecidas à rede bayesiana.

Conforme Murphy (1999):

$$\lambda(x) = P(e_x^-|x)$$

e:

$$\pi(x) = P(x|e_x^+)$$

Considerando a estrutura de poliárvore, que consiste de um nó X , o conjunto de seus pais $\mathbf{U} = \{U_1, \dots, U_n\}$, e o conjunto de seus filhos $\mathbf{Y} = \{Y_1, \dots, Y_m\}$, o conjunto de evidência e_x^+ pode ser decomposto em p subconjuntos disjuntos, um para cada pai de X (CASTILHO; GUTIÉRREZ; HADI, 1998):

$$e_x^+ = \{e_{U_1X}^+, \dots, e_{U_nX}^+\}$$

Onde a evidência $e_{U_jX}^+$ é o subconjunto de e_x^+ contido no subgrafo associado ao nó U_j quando se elimina o arco $U_j \rightarrow X$. Similarmente, o conjunto e_x^- pode ser subdividido em c subconjuntos disjuntos, um para cada filho de X (CASTILHO; GUTIÉRREZ; HADI, 1998):

$$e_x^- = \{e_{XY_1}^-, \dots, e_{XY_m}^-\}$$

Onde a evidência $e_{XY_j}^-$ é o subconjunto de e_x^- contido no subgrafo associado ao nó Y_j quando se elimina o arco $X \rightarrow Y_j$.

Na Figura 4, por exemplo, considerando o conjunto de evidências $e = \{+f, +g, -i\}$, o conjunto $e_E^+ = \{+f\}$ e $e_E^- = \{+g, -i\}$. Do mesmo modo $e_H^+ = \{+f, +g, -i\}$ e $e_H^- = \emptyset$. A eliminação do arco $E \rightarrow H$ divide a rede em dois subgrafos. Neste caso, os conjuntos obtidos são: $e_{EH}^+ = \{+f, -i\}$ e $e_{EH}^- = \{+g\}$.

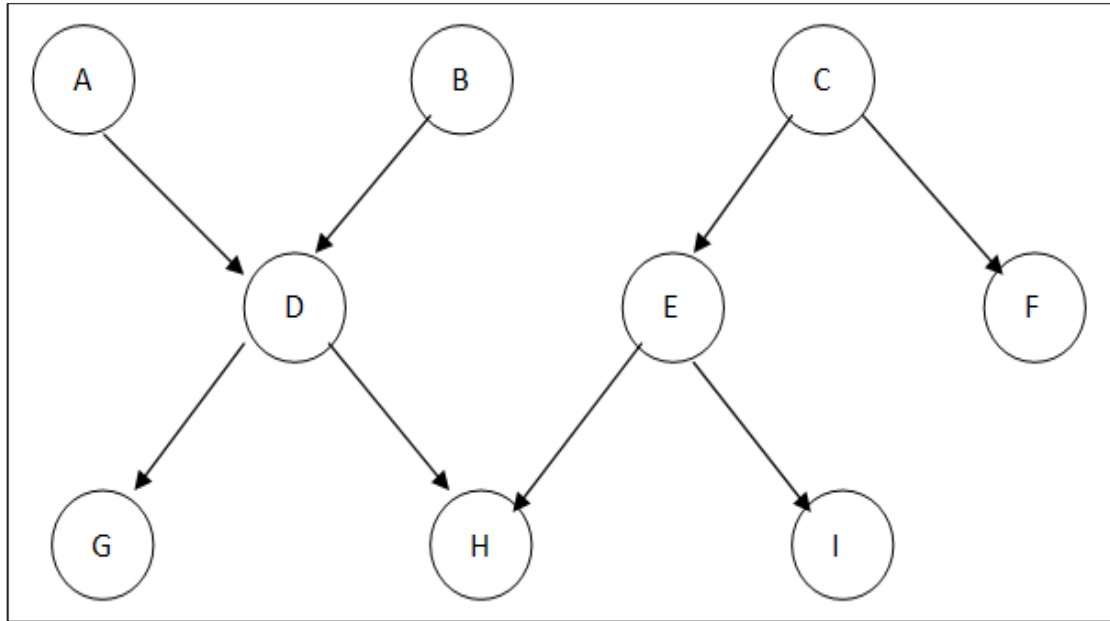


Figura 4. Uma poliárvore
Fonte: DIEZ, F. (2003)

As derivações subseqüentes são baseadas em (PEARL, 1988). Continuando a derivação, a mensagem $\lambda(x)$ pode ser obtida da seguinte forma:

$$\lambda(x) \triangleq P(e_x^- | x)$$

$$\lambda(x) = P(e_{\bar{X}Y_1, \dots, \bar{X}Y_m}^- | x)$$

Os termos $e_{\bar{X}Y_1, \dots, \bar{X}Y_m}^-$ são condicionalmente independentes dado x , portanto, explorando relacionamentos de independência condicional obtém-se:

$$\lambda(x) = P(e_{\bar{X}Y_1}^- | x) P(e_{\bar{X}Y_2}^- | x) \dots P(e_{\bar{X}Y_m}^- | x)$$

A partir da equação acima, é realizada uma generalização destas multiplicações utilizando o símbolo matemático que representa a execução de n destas operações.

$$\lambda(x) = \prod_{j=1}^m \lambda_{Y_j}(x) \quad (5)$$

Onde:

$$\lambda_{Y_j}(x) = P(e_{\bar{X}Y_j}^- | x)$$

E a mensagem $\pi(x)$, obtida como:

$$\pi(x) \triangleq P(x|e_x^+)$$

$$\pi(x) = P(x|e_{U_1X}^+, \dots, e_{U_nX}^+)$$

Utilizando o teorema da probabilidade total condicionada e considerando as relações de independência entre u_n e $e_{U_nX}^+$, onde $n=1\dots n$, obtém-se os somatórios:

$$\pi(x) = \sum_{u_1, \dots, u_n} P(x|u_1, \dots, u_n)P(u_1, \dots, u_n|e_{U_1X}^+, \dots, e_{U_nX}^+)$$

$$\pi(x) = \sum_{u_1, \dots, u_n} P(x|u_1, \dots, u_n)P(u_1|e_{U_1X}^+)P(u_2|e_{U_2X}^+) \dots P(u_n|e_{U_nX}^+)$$

Devido cada par $\{U_i, e_{U_iX}^+\}$ ser independente de $\{U_k, e_{U_kX}^+\}$ para $i \neq k$, considerando:

$$\pi_X(u_i) = P(u_i|e_{U_iX}^+)$$

Então:

$$\pi(x) = \sum_{u_1, \dots, u_n} P(x|u_1, \dots, u_n)\pi_X(u_1)\pi_X(u_2) \dots \pi_X(u_n)$$

$$\pi(x) = \sum_{\mathbf{u}} P(x|\mathbf{u}) \prod_{i=1}^n \pi_X(u_i) \quad (6)$$

Substituindo as equações (5) e (6) na equação (4), obtém-se a equação para determinação das crenças de um nó X qualquer, que consiste em multiplicar as mensagens $\lambda_{Y_j}(x)$ pela multiplicação das mensagens $\pi_X(u_i)$ bem como multiplicá-las pela tabela de probabilidades do nó, identificada por $P(x|\mathbf{u})$, realizar o somatório associado e normalizar, utilizando a constante α :

$$BEL(x) = \alpha \left[\prod_{j=1}^m \lambda_{Y_j}(x) \right] \left[\sum_{\mathbf{u}} P(x|\mathbf{u}) \prod_{i=1}^n \pi_X(u_i) \right]$$

Portanto, o nó X pode calcular suas próprias crenças quando receber as mensagens $\lambda_{Y_j}(x)$ de seus filhos e as mensagens $\pi_X(u_i)$ de seus pais.

Para prescrever como a influência de nova informação se propaga pela rede, é necessário especificar como um nó típico X calcula suas mensagens de saída $\lambda_X(u_i)$, $i = 1, \dots, n$, e $\pi_{Y_j}(x)$, $j = 1, \dots, m$, a partir das mensagens entrantes $\lambda_{Y_j}(x)$ e $\pi_X(u_i)$.

Conforme descrito na equação (4), a atualização das probabilidades em função do fornecimento de determinada evidência é calculada por meio de λ e π . Portanto, para a efetiva propagação das probabilidades calculadas neste contexto, é necessário alterar os valores tanto de λ quanto π , os quais são responsáveis por combinar as mensagens $\lambda_{Y_j}(x)$ e $\pi_X(u_i)$ provenientes dos filhos e pais, respectivamente, em um valor a partir do qual as probabilidades do modelo bayesiano podem ser atualizadas. Portanto, é necessário derivar as equações que calculam estas mensagens.

5.1.1.1 Atualizando λ

Considerando a mensagem $\lambda_X(u_i)$, que o nó X deve enviar para seus pais U_i .

$$\lambda_X(u_i) = P(e_{\bar{U}_i X}^- | u_i) \quad (7)$$

Na derivação de $\lambda_X(u_i)$ é conveniente tratar temporariamente todos os pais (exceto U_i) como uma única variável composta,

$$V = \mathbf{U} - U_i = \{U_1, \dots, U_{i-1}, U_{i+1}, \dots, U_n\} \quad (8)$$

Conectada a X via uma única conexão $V \rightarrow X$, conforme ilustra a Figura 5.

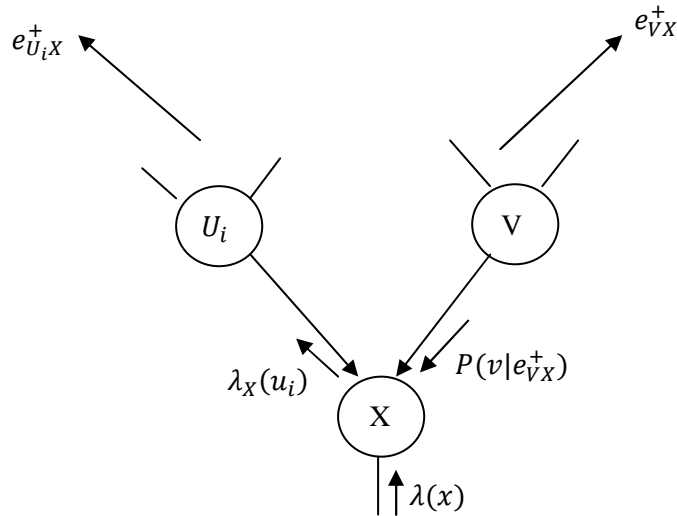


Figura 5. Variáveis envolvidas na atualização de λ
 Fonte: Adaptado de PEARL, J. (1988)

A evidência $e_{U_i X}^-$ governando $\lambda_X(u_i)$ pode ser decomposta em dois componentes:

$$e_{U_i X}^- = \{e_{V X}^+, e_X^-\} \quad (9)$$

Onde:

$$e_{V X}^+ = \bigcup_{k \neq i} e_{U_k X}^+ \quad (10)$$

Substituindo a equação (9) na equação (7), obtém-se:

$$\lambda_X(u_i) = P(e_{V X}^+, e_X^- | u_i)$$

Condicionando em relação a x e v obtém-se os somatórios variando em função das variáveis x e v :

$$\lambda_X(u_i) = \sum_x \sum_v P(e_{V X}^+, e_X^- | u_i, v, x) P(v, x | u_i)$$

Desde que X separa $e_{V X}^+$ de e_X^- e V separa $e_{V X}^+$ de U_i , obtêm-se os seguintes somatórios, nos quais seus termos são obtidos explorando os relacionamentos de

independência condicional entre as variáveis:

$$\lambda_X(u_i) = \sum_x \sum_v P(e_X^-|x)P(e_{VX}^+|v)P(v, x|u_i)$$

Aplicando o teorema de Bayes, Equação (1), ao termo $P(e_{VX}^+|v)$ da equação anterior, definindo a constante $\beta = e_{VX}^+$, explorando propriedades de independência condicional e probabilidade condicional, obtém-se a sucessão de somatórios abaixo:

$$\lambda_X(u_i) = \beta \sum_x \sum_v P(e_X^-|x) \frac{P(v|e_{VX}^+)}{P(v)} P(v, x|u_i)$$

$$\lambda_X(u_i) = \beta \sum_x \sum_v P(e_X^-|x) \frac{P(v|e_{VX}^+) \frac{P(v, x, u_i)}{P(u_i)}}{P(v)}$$

$$\lambda_X(u_i) = \beta \sum_x \sum_v P(e_X^-|x) \frac{P(v|e_{VX}^+) \frac{P(x|v, u_i)P(v, u_i)}{P(u_i)}}{P(v)}$$

$$\lambda_X(u_i) = \beta \sum_x \sum_v P(e_X^-|x) \frac{P(v|e_{VX}^+)}{P(v)} P(x|v, u_i)P(v|u_i)$$

Como v e u_i são independentes, o termo $P(v|u_i)$ da equação anterior torna-se $P(v)$:

$$\lambda_X(u_i) = \beta \sum_x \sum_v P(e_X^-|x) \frac{P(v|e_{VX}^+)}{P(v)} P(x|v, u_i) \cancel{P(v)}$$

Portanto, uma vez que os termos $P(v)$ foram eliminados, obtém-se os somatórios abaixo, cujos valores iterativamente percorrem os valores das variáveis x e y .

$$\lambda_X(u_i) = \beta \sum_x \sum_v P(e_X^-|x)P(v|e_{VX}^+)P(x|v, u_i)$$

Restaurando o significado de V a partir das equações (8) e (10), tem-se:

$$P(x|v, u_i) = P(x|\mathbf{u})$$

$$P(v|e_{VX}^+) = \prod_{k \neq i} P(u_k | e_{VX}^+) = \prod_{k \neq i} P(u_k | e_{U_{kX}}^+) = \prod_{k \neq i} \pi_X(u_k)$$

e $\lambda_X(u_i)$ se torna a equação abaixo, utilizada dentro do contexto de propagação de mensagens entre os nós para enviar uma mensagem do nó filho para seus respectivos pais:

$$\lambda_X(u_i) = \beta \sum_x \lambda(x) \sum_{u_k: k \neq i} P(x|\mathbf{u}) \prod_{k \neq i} \pi_X(u_k)$$

5.1.1.2 Atualizando π

Considerando a mensagem $\pi_{Y_j}(x)$, que o nó X envia para seus filhos Y_j .

$$\pi_{Y_j}(x) = P(x|e_{XY_j}^+)$$

$e_{XY_j}^+$ significa o conjunto de evidência completo E, excluindo a evidência encontrada na sub-rede associada ao nó Y_j ,

$$e_{XY_j}^+ = \mathbf{e} - e_{XY_j}^-$$

Logo, $\pi_{Y_j}(x)$ é igual a BEL(x) quando a evidência $e_{XY_j}^-$ é suprimida.

Equivalentemente, a expressão $\pi_{Y_j}(x)$ pode ser obtida a partir de BEL(x), configurando

$\lambda_{Y_j}(x) = 1$ obtendo desta forma, a seguinte equação:

$$\pi_{Y_j}(x) = \alpha \prod_{k \neq j} \lambda_{Y_k}(x) \pi(x)$$

Alternativamente, $\pi_{Y_j}(x)$ pode ser obtida a partir de BEL(x) escrevendo:

$$\pi_{Y_j}(x) = \alpha \frac{BEL(x)}{\lambda_{Y_j}(x)}$$

5.1.1.3 Resumo do algoritmo de propagação em poliárvores

O resumo abaixo descrito é baseado em (KORB; NICHOLSON, 2004). Considerando uma variável X , com m filhos e n pais. A distribuição de probabilidades da variável X pode ser calculada se três tipos de parâmetros estão disponíveis:

- a) A força atual do suporte causal, π , formado pela contribuição de cada arco de chegada $U_i \rightarrow X$:

$$\pi_X(x_i) = P(u_i | e_{U_i X}^+)$$

- b) A força atual do suporte diagnóstico, λ , formado pela contribuição de arco de saída $X \rightarrow Y_j$:

$$\lambda_{Y_j}(x) = P(e_{XY_j}^- | x)$$

- c) A matriz de probabilidade condicional fixa $P(x | u_1, \dots, u_n)$ que relaciona a variável X a seus pais imediatos.

Utilizando estes parâmetros, a atualização de crença local pode ser executada em três passos:

Passo 1- Atualização de crença: quando um nó X é ativado (recebe uma evidência), ele simultaneamente inspeciona as mensagens $\pi_X(x_i)$ enviadas pelos seus pais e as mensagens $\lambda_{Y_j}(x)$ enviadas pelos seus filhos. Utilizando esta entrada, o nó atualiza sua medida de crença para:

$$BEL(x) = \alpha \lambda(x) \pi(x)$$

Onde:

$$\lambda(x) = \prod_{j=1}^m \lambda_{Y_j}(x)$$

$$\pi(x) = \sum_{\mathbf{u}} P(x | \mathbf{u}) \prod_{i=1}^n \pi_X(u_i)$$

Passo 2- Propagação ascendente: Utilizando as mensagens recebidas, o nó X calcula novas mensagens λ para serem enviadas aos seus pais. Por exemplo, a nova mensagem $\lambda_X(u_i)$ que X envia a seus pais U_i é calculada por:

$$\lambda_X(u_i) = \beta \sum_x \lambda(x) \sum_{u_k: k \neq i} P(x|\mathbf{u}) \prod_{k \neq i} \pi_X(u_k)$$

Passo 3- Propagação descendente: Cada nó calcula novas mensagens π para serem enviadas aos seus filhos. Por exemplo, a nova mensagem $\pi_{Y_j}(x)$ que X envia aos seus filhos Y_j é calculada por:

$$\pi_{Y_j}(x) = \alpha \left[\prod_{k \neq j} \lambda_{Y_k}(x) \right] \sum_{\mathbf{u}} P(x|\mathbf{u}) \prod_{i=1}^n \pi_X(u_i)$$

$$\pi_{Y_j}(x) = \alpha \frac{BEL(x)}{\lambda_{Y_j}(x)}$$

Utilizando os três passos acima descritos é possível realizar inferência probabilística em poliárvores de uma forma computacionalmente eficiente. Considerando que os algoritmos de inferência foram criados a alguns anos e devido a necessidade de sua utilização para representação de problemas a partir de redes bayesianas, eles foram desenvolvidos em diversas ferramentas, livres ou proprietárias, para que as pessoas envolvidas na modelagem de uma rede bayesiana dediquem-se a isto e não lidem com detalhes de implementação dos diferentes algoritmos existentes.

6 TRABALHOS CORRELATOS

Este capítulo apresenta um conjunto de ferramentas que utilizam redes bayesianas como abordagem para raciocínio sob incerteza e disponibilizam um ou mais algoritmos de inferência.

O desenvolvimento de ferramentas para redes bayesianas tem por objetivo facilitar a modelagem e resolução de problemas que apresentam características aplicáveis a esta teoria. Neste sentido, diversos softwares foram desenvolvidos, disponibilizados ou comercializados. Esta seção apresenta algumas destas *shells* bem como algumas de suas características cujas descrições são baseadas em (KORB; NICHOLSON, 2004).

6.1 ANALYTICA

A ferramenta Analytica foi desenvolvida pela empresa Lumina Decision Systems Inc fundada em 1991 por Max Henrion e Brian Arnold. A ênfase deste software consiste na utilização de diagramas de influência como uma ferramenta de suporte para decisão estatística. No entanto, não utiliza a terminologia de redes bayesianas, o que pode dificultar a identificação de aspectos de sua funcionalidade. A Figura 5 apresenta a interface gráfica desta ferramenta.

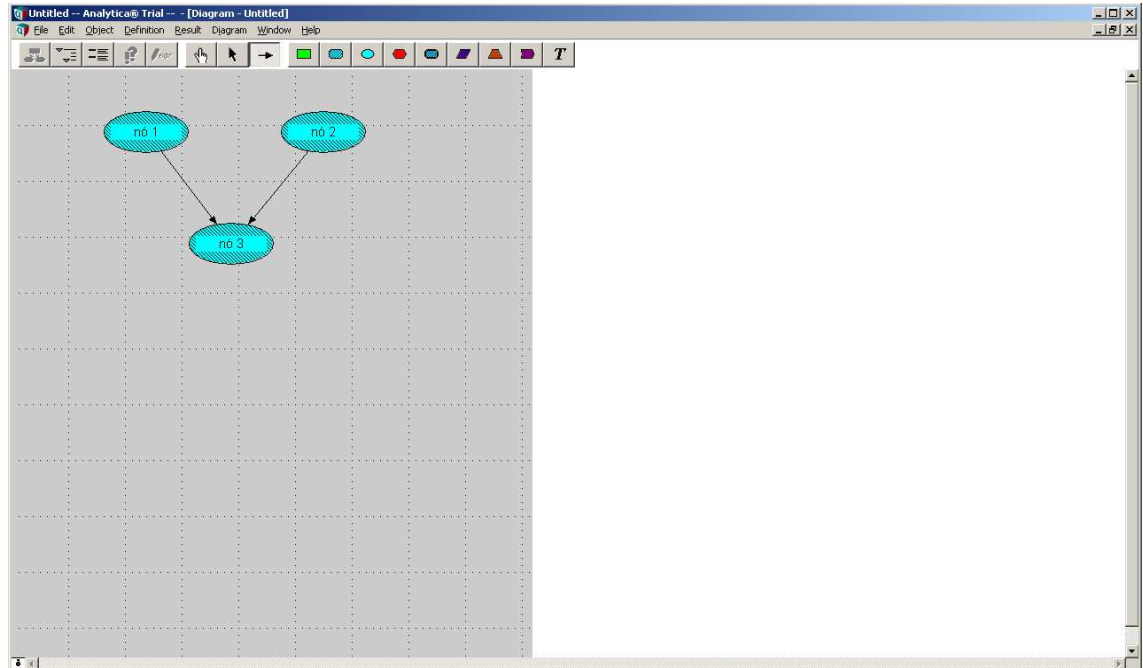


Figura 6. Interface gráfica do Analytica
 Fonte: Lumina Decision Systems (1991)

Está disponível para Windows e Macintosh, sendo sua Application Programming Interface (API)⁸, denominada Analytica Decision Engine, disponível para os sistemas operacionais Windows 95/98 ou NT 4.0 e suporta diversas distribuições contínuas e discretas, além de fornecer um grande número de funções matemáticas e estatísticas.

Para a inferência, o Analytica fornece amostragem MDMC⁹ básica, entre outros, e permite que o tamanho da amostra seja especificado. A interface gráfica do Analytica fornece muitas formas de visualizar os resultados da inferência por meio de tabelas e gráficos.

Outra característica do Analytica é o suporte a construção de grandes modelos, possibilitando a combinação hierárquica de modelos menores, conectando os nós de entrada e saída especificados.

⁸Conjuntos de bibliotecas pré-definidas que podem ser reutilizadas em diferentes aplicações. (DEITEL, H.; DEITEL, P., 2005)

⁹Gera uma amostra alterando aleatoriamente a amostra anterior. (KORB; NICHOLSON, 2004)

6.2 BAYESIANLAB

O BayesianLab foi desenvolvido pela empresa BAYESIA e sua interface gráfica está disponível para todas as plataformas que suportam Java Runtime Environment (JRE). A Figura 6 apresenta a interface gráfica desta ferramenta.

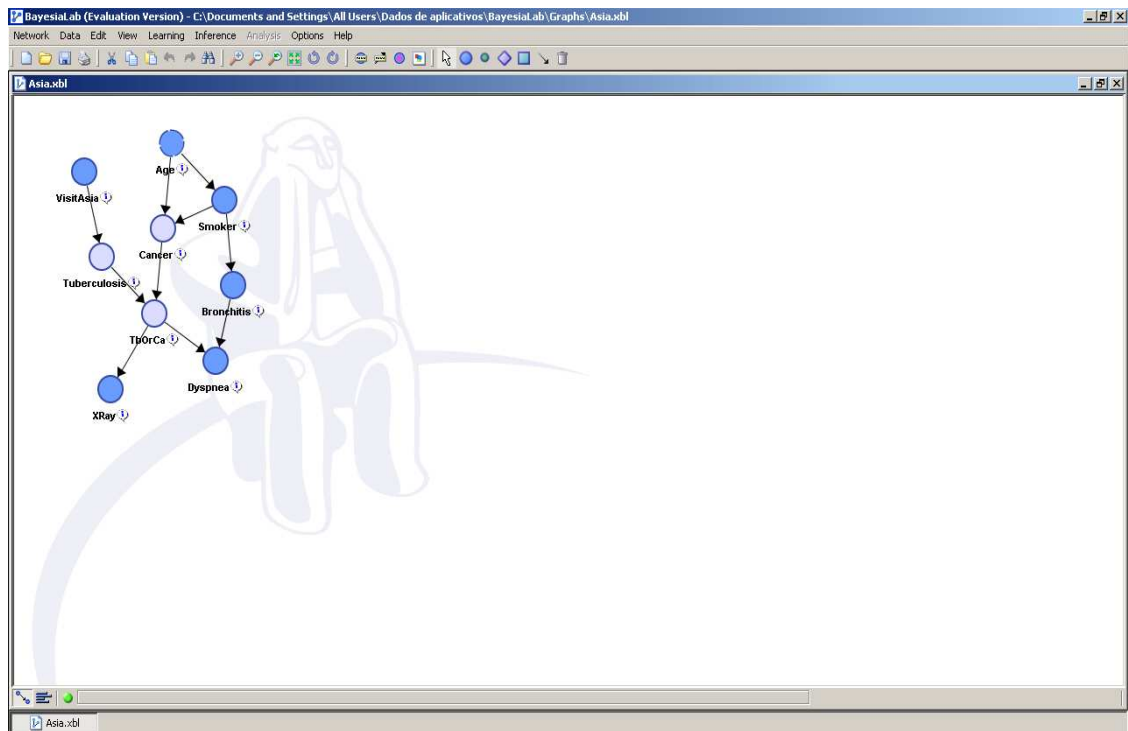


Figura 7. Interface gráfica da *shell* Bayesian
Fonte: BAYESIA (2001)

Nesta ferramenta, variáveis contínuas devem ser discretizadas. Quanto ao aprendizado das variáveis a partir de uma base de dados, a *shell* suporta intervalos de igual distância¹⁰, intervalos de igual frequência¹¹ e uma discretização de árvore de decisão que escolhe os intervalos dependendo da informação que eles contribuem para uma variável alvo especificada.

A inferência no BayesianLab é realizada com o que eles denominam modo

¹⁰ O conjunto é dividido em intervalos do mesmo tamanho (BAYESIA, 2004).

¹¹ Utiliza intervalos com o mesmo número de casos associados (BAYESIA, 2004).

de validação. Um algoritmo Junction Tree¹² é utilizado como padrão possuindo três métodos para aprendizado estrutural: SopLEQ¹³, que utiliza propriedades de equivalência em redes bayesianas e duas versões da busca em Taboo¹⁴.

6.3 GRAPHICAL NETWORK INTERFACE (GENIE)

Desenvolvido pelo grupo de sistemas de decisão de Druzdzel, o suporte desse aplicativo a redes de decisão, além de redes bayesianas, reflete seus interesses de ensino, pesquisa, suporte a decisão e engenharia do conhecimento. A Figura 8 apresenta a interface gráfica do GeNie.

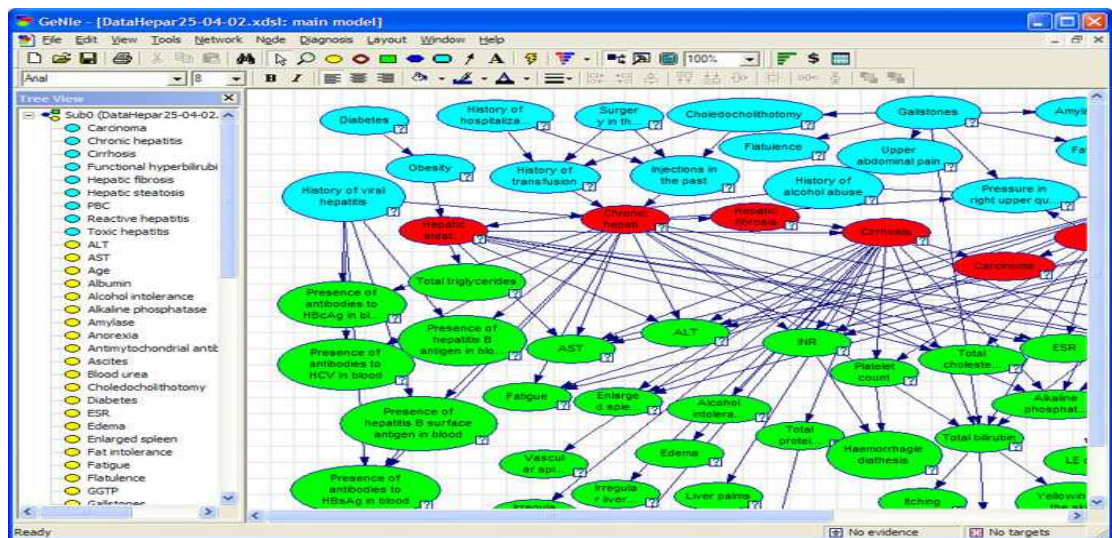


Figura 8. Interface gráfica do GeNie.
Fonte: GeNie, (1998)

O GeNie é um ambiente de desenvolvimento para construir redes de decisão para a plataforma Windows enquanto que o Structural Modeling, Reasoning and

¹² Algoritmo de inferência exata utilizado para propagação de crenças em redes bayesianas (LUNA, 2004)

¹³ Método de busca baseado na caracterização global dos dados e na exploração das propriedades de redes (BAYESIA, 2004).

¹⁴ Este algoritmo mantém uma lista de tabus de k estados visitados anteriormente que não podem ser novamente visitados (RUSSEL; NORVIG, 2004).

Learning Engine (SMILE) é um mecanismo de inferência portátil, que apresenta uma biblioteca de classes C++ correntemente compilada para Windows, Solaris e Linux.

O algoritmo de inferência padrão é o Junction Tree, no entanto, um algoritmo para poliárvores também está disponível além de diversos algoritmos de inferência aproximada, como logic sampling¹⁵, likelihood weighting¹⁶, entre outros.

6.4 HUGIN EXPERT

A *shell* Hugin Expert foi inicialmente desenvolvida por um grupo na Universidade de Aalborg, como parte do projeto ESPRIT. O desenvolvimento do Hugin continuou por meio de outro projeto de Lauritzen-Jensen chamado ODIN. A estreita conexão entre o Hugin Expert e o grupo de pesquisa em Aalborg continuou, incluindo co-localização e movimentação de pessoal entre os dois.

Isto significa que o Hugin Expert contribuiu consistentemente e tomou vantagem das últimas pesquisas em redes bayesianas devido à interação direta com o ambiente acadêmico de pesquisa. Em 1998 a Hewlett-Packard comprou 45% da Hugin Expert e estabeleceu uma nova companhia independente chamada Dezide, que baseia seus produtos (“dezisionWorks”) no Hugin. Posteriormente, Hugin é também um revendedor do dezisionWorks sob a denominação de Hugin Advisor. A Figura 8 ilustra a interface gráfica da *shell*.

¹⁵ É o algoritmo de amostragem mais simples. Ele gera um caso selecionando valores aleatoriamente para cada nó, ponderada pela probabilidade deste valor ocorrer (KORB; NICHOLSON, 2004).

¹⁶ É uma modificação do algoritmo logic sampling que resolve o problema com evidência improvável, sempre empregando o valor amostrado para cada nó evidência (KORB; NICHOLSON, 2004).

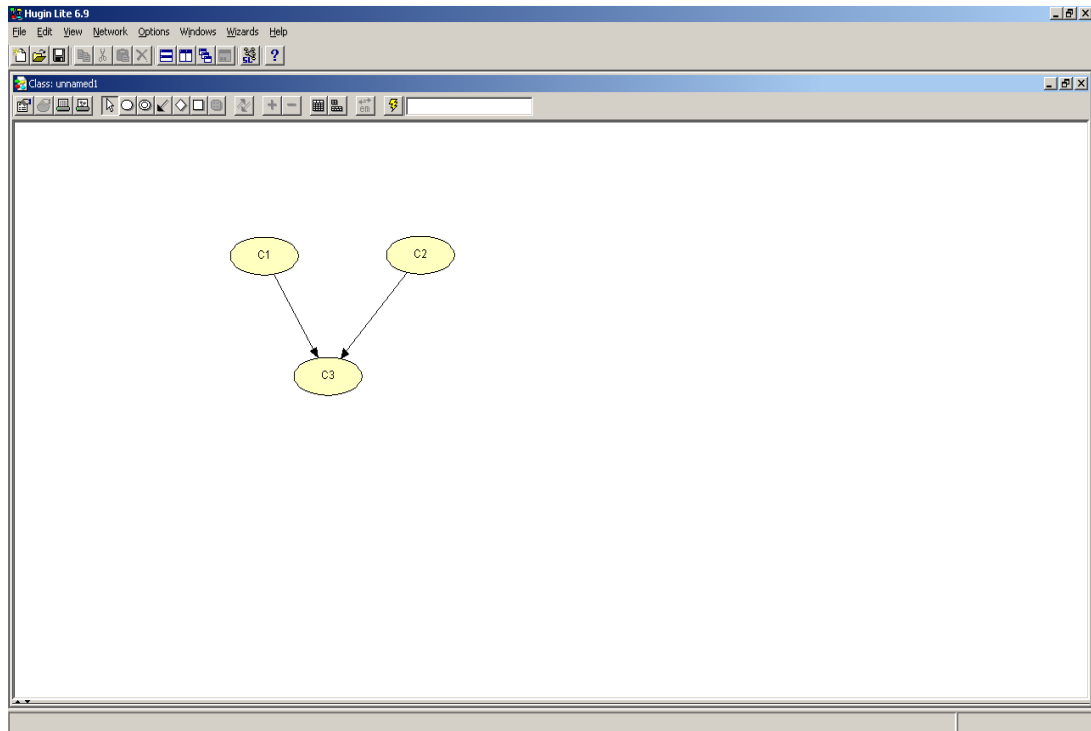


Figura 9. Interface gráfica da ferramenta Hugin Expert
 Fonte: Hugin Expert (1989)

A API do Hugin é chamada de “Hugin Decision Engine” e está disponível para as linguagens C++, Java e como um servidor Active-X podendo ser executada nos sistemas operacionais Sun Solaris (Sparc e x86), HP-UX, Linux e Windows. A interface gráfica do Hugin está disponível para Sun Solaris (Sparc, x86), Windows e Linux Red-hat.

Há suporte para modelagem de variáveis contínuas e combinação de nós discretos e contínuos. As tabelas de probabilidades condicionais podem ser especificadas com expressões assim como entradas manuais, sendo que sua soma tem que ser um. Entradas que não somam um (1) são normalizadas.

A inferência é realizada utilizando o algoritmo Junction Tree com opções para escolher possíveis variações. Há a opção de variar o método de triangulação¹⁷, e

¹⁷ Consiste na introdução de arcos em loops com mais de três nós. (Algoritmos de inferencia em redes probabilísticas baseadas em árvores de junção) Marcelo Ladeira

outra para realizar compressão. Uma versão aproximada do algoritmo da árvore de junção é oferecida, onde todas as probabilidades menores do que determinado limiar são configuradas para zero.

Além do suporte a redes bayesianas, o aplicativo também suporta redes de Decisão. Neste caso, a *shell* requer a existência de um caminho dirigido incluindo todas as variáveis de decisão, ele fornece a utilidade esperada de cada opção de decisão na tabela de decisão.

O aprendizado de parâmetros é realizado com o algoritmo EM o estrutural é realizado utilizando o algoritmo PC, além de fornecer suporte para redes bayesianas orientadas a objeto.

6.5 NETICA

O desenvolvimento da *shell* Netica iniciou em 1992, por Norsys CEO Brent Boerlage, cujo mestrado havia sido recentemente concluído na Universidade British Columbia sendo comercialmente disponibilizada em 1995. A Figura 9 apresenta o ambiente gráfico desta ferramenta.

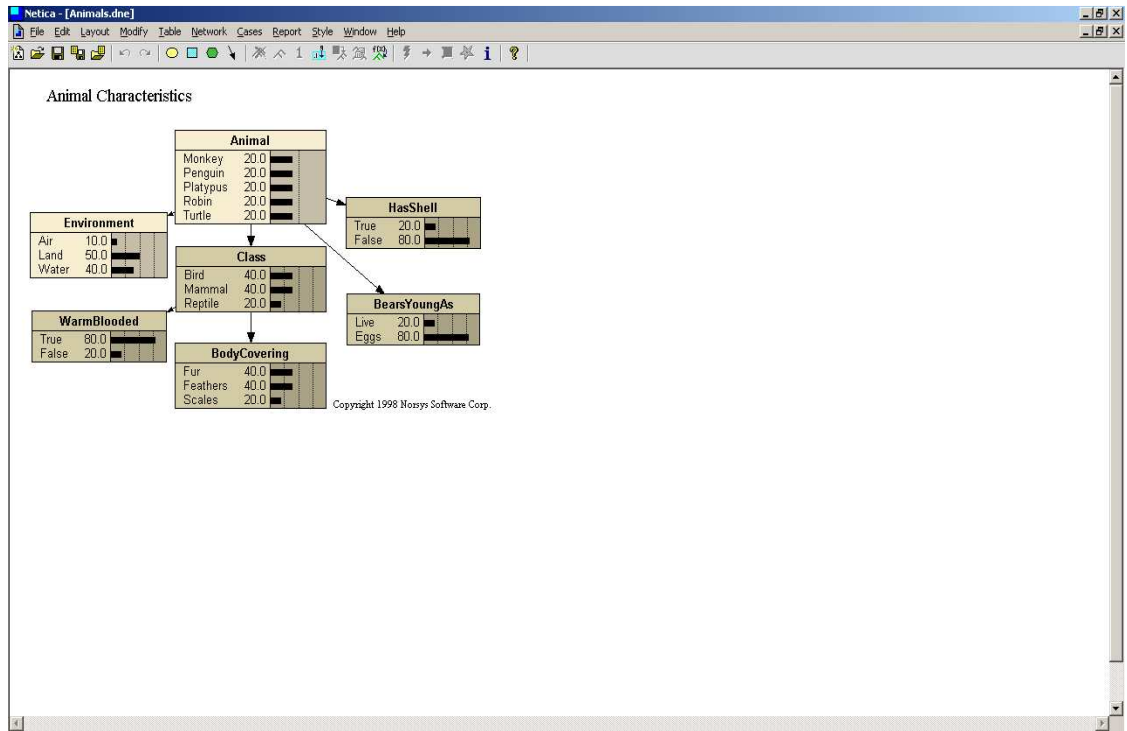


Figura 10. Interface gráfica da *shell* Netica
 Fonte: Norsys Software Corp (1995)

A API da shell Netica está disponível para as linguagens C, C#, VB, C++ e Java, podendo ser utilizada nos sistemas operacionais Mac OSX, Sun Sparc, Linux e Windows e sua interface gráfica está disponível para Mac e Windows sendo seu algoritmo de inferência o junction tree.

7 O FORMALISMO BAYESIANO NA SHELL PEGASUS UNCERTAINTY MODELING

Esta pesquisa é parte integrante do desenvolvimento de uma ferramenta para modelagem da incerteza, que utiliza os fatores de certeza, a lógica *fuzzy*, a teoria de Dempster-Shafer e redes bayesianas, a partir das quais é possível modelar diferentes problemas baseando-se em suas características inerentes. Diversas ferramentas, tanto livres quanto proprietárias, disponibilizam estas teorias, freqüentemente de forma isolada para facilitar o processo de modelagem e resolução.

No entanto, devido à carência de ferramentas que provêm duas ou mais metodologias para modelagem da incerteza, em um mesmo ambiente de desenvolvimento, esta pesquisa resultou no início da criação da *shell* Pegasus *Uncertainty Modeling*, cuja funcionalidade abrange a disponibilização destas teorias em um ambiente integrado, propiciando um maior número de alternativas para a solução de problemas em potencial.

Neste sentido, este trabalho corresponde ao desenvolvimento da teoria de redes bayesianas na *shell* Pegasus, cujo objetivo é modelar relacionamentos probabilísticos entre as variáveis que constituem o problema e fornecer um mecanismo de apoio a decisão em diferentes áreas.

Considerando especificamente as redes bayesianas, esta pesquisa abordou o problema da inferência por meio de um algoritmo de passagem de mensagens em redes cuja estrutura é caracterizada por uma poliárvore (PEARL, 1988), e assim, foi realizado um levantamento a partir de bibliografias nacionais e internacionais tornando possível a obtenção dos resultados desta pesquisa.

Este capítulo apresenta detalhadamente a implementação e os resultados obtidos, além de descrever as etapas metodológicas, as quais consistem em um

levantamento bibliográfico, na seleção e estudo de algumas *shells* existentes, na demonstração matemática do algoritmo de inferência implementado, bem como a modelagem tanto da *shell* Pegasus quanto do módulo de redes bayesianas, a partir da qual a implementação foi efetivada e em seguida foram desenvolvidos testes com o intuito de avaliar os resultados produzidos pelo mecanismo de inferência.

7.1 ESTUDO DAS *SHELLS* DISPONÍVEIS

Após o levantamento bibliográfico, foi realizado o levantamento de algumas ferramentas utilizadas para inferência bayesiana disponíveis com o intuito de definir os algoritmos e padrões de interface que elas utilizam e, portanto, desenvolver o módulo bayesiano da *shell* Pegasus segundo características comuns destas ferramentas. Concluída esta etapa, foi realizado um exemplo numérico da utilização do algoritmo de inferência.

7.2 DEMONSTRAÇÃO MATEMÁTICA DO MECANISMO DE INFERÊNCIA

Primeiramente, o pseudocódigo (NEOPOLITAN, 2004), segundo o qual esta implementação foi baseada é descrito, seguindo-se com a apresentação de um exemplo numérico que demonstra a utilização do algoritmo. Este pseudocódigo é constituído por quatro métodos, cuja nomenclatura utilizada é a mesma da bibliografia base (NEOPOLITAN, 2004): *initial_net()*, *update_tree()*, *send_λ_msg()* e *send_π_msg()*. Os parâmetros destes métodos foram adaptados com o intuito de refletir a implementação realizada.

O método *initial_net()*, ilustrado na Figura 11, é utilizado para inicializar os valores π , que é o produto da tabela de probabilidade do respectivo nó pelas mensagens $\pi_Y(x)$ proveniente de seus pais e λ , representa o produto das mensagens $\lambda_X(z)$ proveniente de seus filhos se nenhuma evidência é fornecida para X (KORB; NICHOLSON, 2004), bem como as mensagens $\lambda_X(z)$ e $\pi_Y(x)$ que um nó filho envia para seus pais, e que um pai envia para seus filhos, respectivamente, utilizados no procedimento de inferência sendo que foi escolhido devido ao fácil acesso a sua documentação e devido sua importância histórica para a área de redes bayesianas.

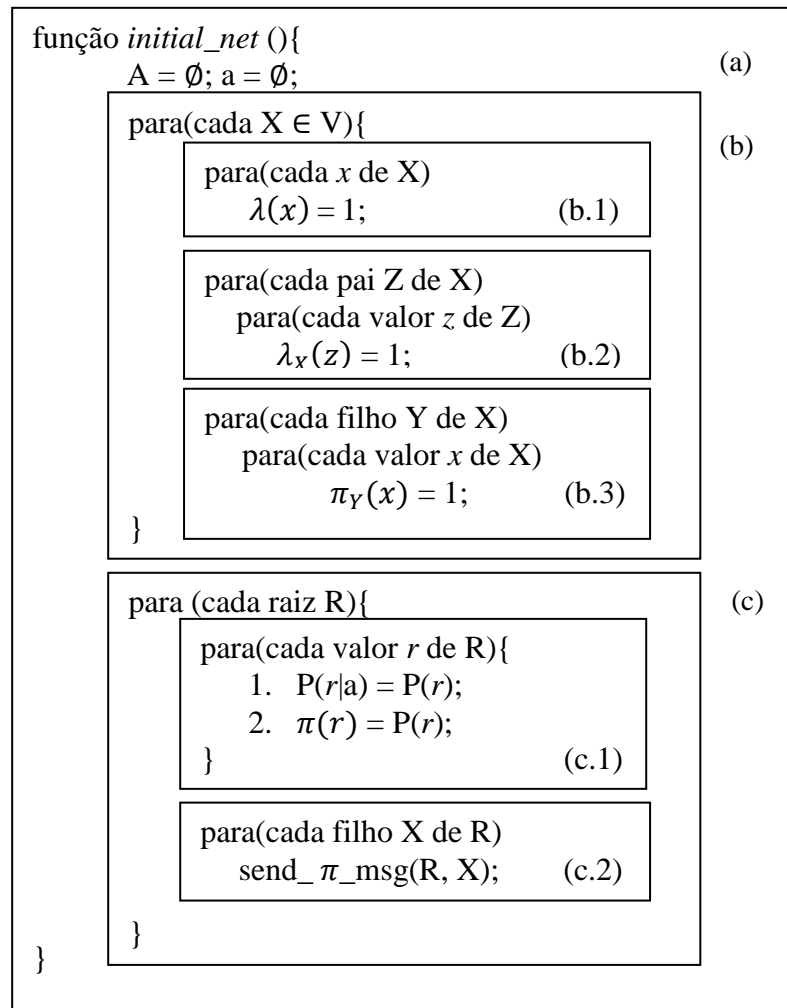


Figura 11. Método *initial_net()*

Fonte: Adaptado de NEOPOLITAN, R. (2004)

O método ilustrado na Figura 11 fornece os valores iniciais necessários tanto para a inicialização do modelo quanto para a atualização da distribuição de probabilidades quando uma evidência é fornecida, sendo esta última tarefa delegada ao procedimento *update_tree()*, ilustrado na Figura 12, que recebe como argumentos uma variável V , que caracteriza o nó evidência, e o estado ao qual atribui-se a evidência.

```

função update_tree (variável  $V$ , valor-da-variável  $^v$ ){
  1.  $A = A \cup \{V\}$ ;  $a = a \cup \{^v\}$ ;
  2.  $\lambda(^v) = 1$ ;  $\pi(^v) = 1$ ;  $P(^v|a) = 1$ ;
  3. para(cada valor de  $v \neq ^v$ )
  4.    $\lambda(v) = 0$ ;  $\pi(v) = 0$ ;  $P(v|a) = 0$ ;
  5. para(cada pai  $Z$  de  $V$  tal que  $Z \notin A$ )
  6.   send_λ_msg( $V, Z$ );
  7. para(cada filho  $X$  de  $V$ )
  8.   send_π_msg( $V, X$ );
}

```

Figura 12. Método *update_tree()*
 Fonte: Adaptado de NEOPOLITAN, R. (2004)

O método *send_λ_msg()* ilustrado na Figura 13 é responsável pela propagação da mensagem que um nó filho envia para seus pais, portanto, descrevendo um fluxo de transferência de mensagens ascendente, caracterizando um mecanismo de raciocínio diagnóstico.

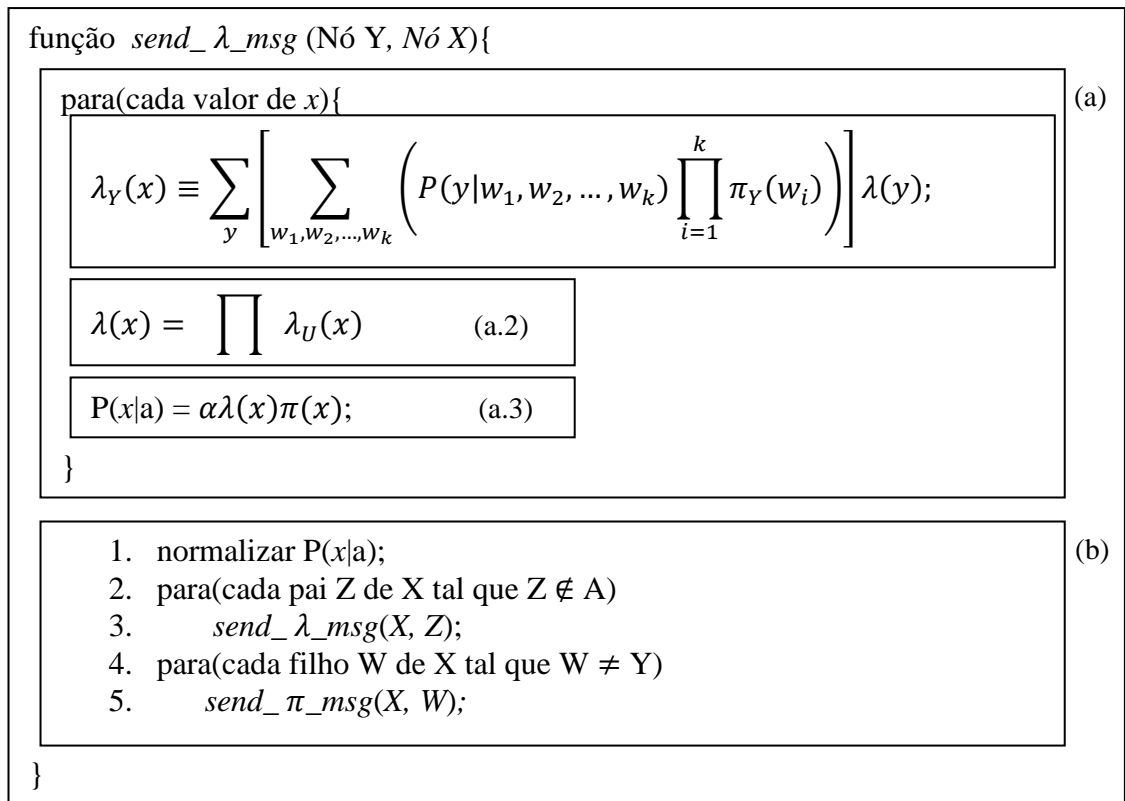


Figura 13. Método *send_λ_msg*()

Fonte: Adaptado de NEOPOLITAN, R. (2004)

Finalizando a descrição do pseudocódigo, o método *send_π_msg*(), ilustrado na Figura 14, é responsável pela transferência de mensagens que um nó pai envia para seus filhos, portanto, descrevendo um fluxo de transferência de mensagens descendente que caracteriza um mecanismo causal para propagação de probabilidades na rede bayesiana.

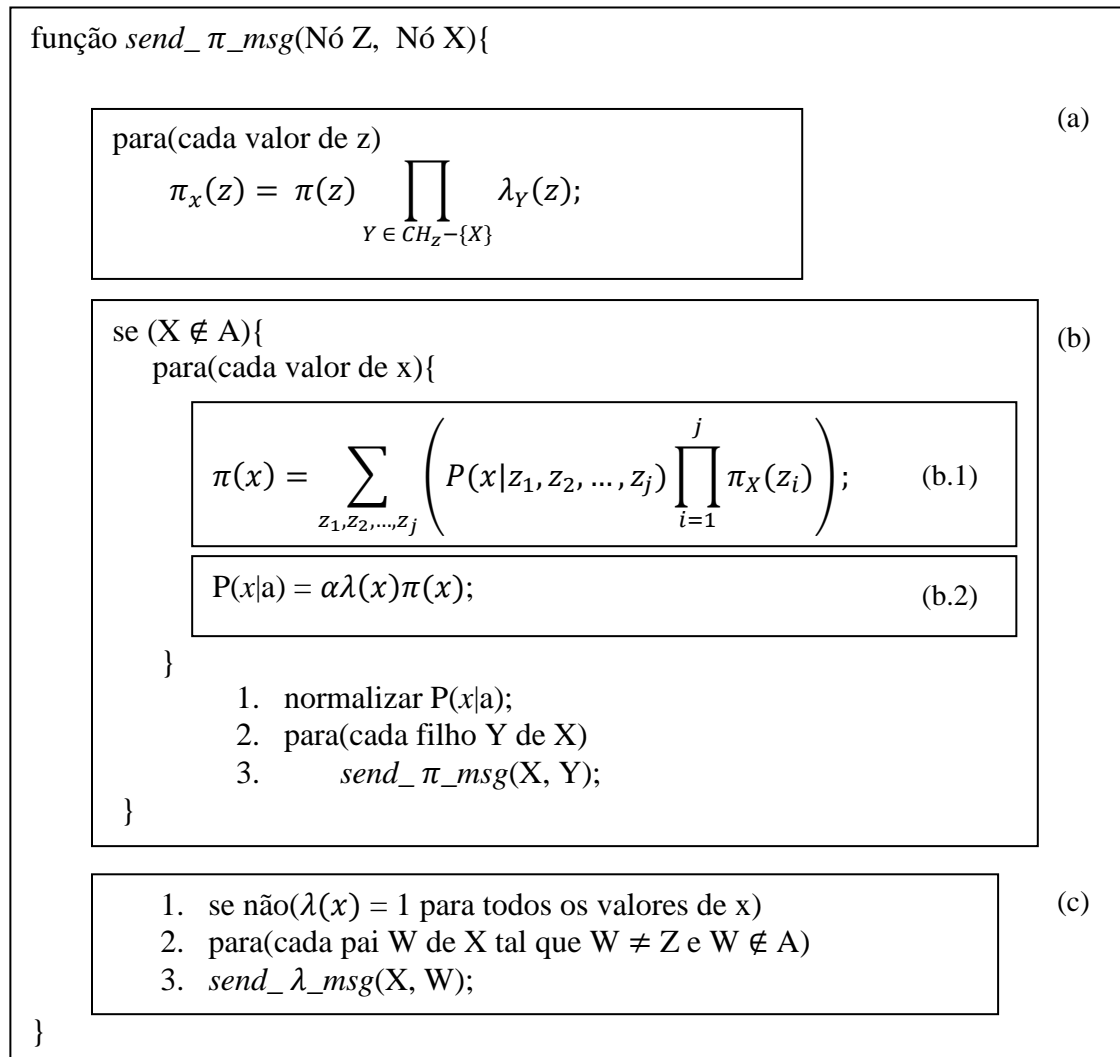


Figura 14. Método *send_π_msg*

Fonte: Adaptado de NEOPOLITAN, R. (2004)

Para demonstrar o funcionamento do algoritmo de inferência será utilizada uma rede bayesiana composta de três nós responsável pelo diagnóstico de gripe. Os nós representam as variáveis Febre, Dor de Cabeça e Gripe e considera-se que cada nó possui apenas dois estados. Os valores associados às tabelas de probabilidade da Figura 15 são hipotéticos, utilizados somente para exemplificação. A Figura 15 ilustra o modelo criado para o exemplo e a Tabela 4 apresenta o nome dos nós, seus respectivos estados e significados.

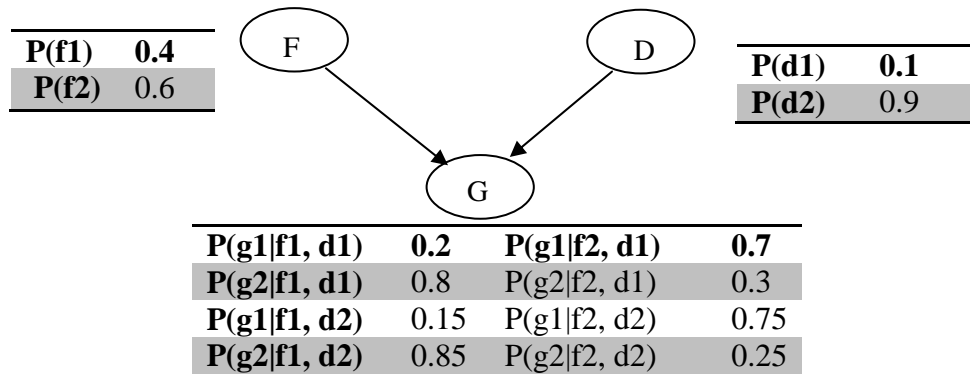


Figura 15. Rede usada no exemplo

Tabela 4. Variáveis, estados e seus significados

F	f1	Possui febre
	f2	Não possui febre
D	d1	Está com dor no corpo
	d2	Não está com dor no corpo
G	g1	Tem gripe
	g2	Não tem gripe

Primeiramente, a rede bayesiana deve ser inicializada. Isto ocorre realizando uma chamada ao método *initial_tree()*, previamente descrito no pseudocódigo, resultando nos seguintes valores:

$$A = \emptyset;$$

$$a = \emptyset;$$

As linhas acima correspondem a execução da parte (a) do pseudocódigo *initial_tree()*.

$$\lambda(f1) = 1; \lambda(f2) = 1;$$

$$\lambda(d1) = 1; \lambda(d2) = 1;$$

$$\lambda(g1) = 1; \lambda(g2) = 1;$$

As linhas acima referem-se a parte (b.1) do método *initial_tree()*, onde são definidos os valores $\lambda()$ como 1 para todos os nós, Febre (F), Dor de cabeça (D) e Gripe (G), e estados correspondentes, da rede bayesiana.

$$\lambda_G(f1) = 1; \lambda_G(f2) = 1;$$

$$\lambda_G(d1) = 1; \lambda_G(d2) = 1;$$

Execução da parte (b.2), na qual se define as mensagens que o nó G envia para seus pais F e D, $\lambda_G()$ com o valor 1.

$$\begin{aligned} \pi_G(f1) &= 1; \pi_G(f2) = 1; \\ \pi_G(d1) &= 1; \pi_G(d2) = 1; \end{aligned}$$

Execução da parte (b.3), define as mensagens $\pi_G()$ com o valor 1.

$$\begin{aligned} P(f1|\emptyset) &= P(f1) = 0.4; \\ P(f2|\emptyset) &= P(f2) = 0.6; \end{aligned}$$

Execução da parte (c.1), linha 1, inicializa as probabilidades condicionais do nó F, como o termo condicionante é \emptyset , assume a probabilidade *a priori*.

$$\begin{aligned} \pi(f1) &= P(f1) = 0.4; \\ \pi(f2) &= P(f2) = 0.6; \end{aligned}$$

Parte (c.1) linha 2, inicializa os valores de $\pi(F)$

send_ π _msg(F, G);

Executa uma chamada a send_ π _msg(F, G), correspondente a parte (c.2) do pseudocódigo.

$$\begin{aligned} P(d1|\emptyset) &= P(d1) = 0.1; \\ P(d2|\emptyset) &= P(d2) = 0.9; \end{aligned}$$

Execução da parte (c.1) linha 1, do pseudocódigo.

$$\begin{aligned} \pi(d1) &= P(d1) = 0.1; \\ \pi(d2) &= P(d2) = 0.9; \end{aligned}$$

Execução da parte parte (c.1) linha 2 do pseudocódigo.

send_ π _msg(D, G);

A chamada send_ π _msg(F, G) resulta nos seguintes valores:

$$\pi_G(f1) = \pi(f1) = 0.4;$$

$$\pi_G(f2) = \pi(f2) = 0.6;$$

Execução da parte (a) do pseudocódigo.

$$\begin{aligned}\pi(g1) &= P(g1|f1, d1)\pi_G(f1)\pi_G(d1) + P(g1|f1, d2)\pi_G(f1)\pi_G(d2) \\ &+ P(g1|f2, d1)\pi_G(f2)\pi_G(d1) + P(g1|f2, d2)\pi_G(f2)\pi_G(d2) \\ \pi(g1) &= (0.2)(0.4)(1) + (0.15)(0.4)(1) + (0.7)(0.6)(1) + (0.75)(0.6)(1) \\ \pi(g1) &= 1.01\end{aligned}$$

Parte (b.1), corresponde ao somatório das multiplicações dos valores da tabela de probabilidade pelas mensagens $\pi_X(z_i)$ que o filho G envia para seus pais F e D, no estado g1.

$$\begin{aligned}\pi(g2) &= P(g2|f1, d1)\pi_G(f1)\pi_G(d1) + P(g2|f1, d2)\pi_G(f1)\pi_G(d2) \\ &+ P(g2|f2, d1)\pi_G(f2)\pi_G(d1) + P(g2|f2, d2)\pi_G(f2)\pi_G(d2) \\ \pi(g2) &= (0.8)(0.4)(1) + (0.85)(0.4)(1) + (0.3)(0.6)(1) + (0.25)(0.6)(1) \\ \pi(g2) &= 0.99\end{aligned}$$

Parte (b.1), percorrendo o segundo estado do nó G, que corresponde ao somatório das multiplicações dos valores da tabela de probabilidade pelas mensagens $\pi_X(z_i)$ que o filho G envia para seus pais F e D.

$$\begin{aligned}P(g1|\emptyset) &= \alpha\lambda(g1)\pi(g1) = \alpha(1)(1.01) = 1.01\alpha \\ P(g2|\emptyset) &= \alpha\lambda(g2)\pi(g2) = \alpha(1)(0.99) = 0.99\alpha\end{aligned}$$

Parte (b.2). Calcula a probabilidade de G quando nenhuma evidência é fornecida.

$$\begin{aligned}P(g1|\emptyset) &= \frac{1.01\alpha}{1.07\alpha + 0.99\alpha} = 0.505 \\ P(g2|\emptyset) &= \frac{0.99\alpha}{1.01\alpha + 0.99\alpha} = 0.495\end{aligned}$$

Parte (b) linha 1. Efetiva a normalização das probabilidades para que a soma das mesmas seja 1.

Após realizada a normalização, o laço *para* do pseudocódigo `send_pi_msg()`, situado na parte (b), linha 2, é executado e o nó D envia uma mensagem para seu filho G, ação esta executada pela chamada `send_pi_msg(D, G)` resultando nos

seguintes valores:

$$\begin{aligned}\pi_G(d1) &= \pi(d1) = 0.1; \\ \pi_G(d2) &= \pi(d2) = 0.9;\end{aligned}$$

Parte (a)

$$\begin{aligned}\pi(g1) &= P(g1|f1, d1)\pi_G(f1)\pi_G(d1) + P(g1|f1, d2)\pi_G(f1)\pi_G(d2) \\ &+ P(g1|f2, d1)\pi_G(f2)\pi_G(d1) + P(g1|f2, d2)\pi_G(f2)\pi_G(d2) \\ \pi(g1) &= (0.2)(0.4)(0.1) + (0.15)(0.4)(0.9) + (0.7)(0.6)(0.1) + \\ &(0.75)(0.6)(0.9) \\ \pi(g1) &= 0.509\end{aligned}$$

Parte (b.1), corresponde ao somatório das multiplicações dos valores da tabela de probabilidade pelas mensagens $\pi_X(z_i)$ que o filho G envia para seus pais F e D, no estado g1. Porém, nesta segunda execução, os valores de $\pi_G(D)$ foram atualizados para 0.1 e 0.9, enquanto que, nos cálculos anteriores seus valores eram 1.

$$\begin{aligned}\pi(g2) &= P(g2|f1, d1)\pi_G(f1)\pi_G(d1) + P(g2|f1, d2)\pi_G(f1)\pi_G(d2) \\ &+ P(g2|f2, d1)\pi_G(f2)\pi_G(d1) + P(g2|f2, d2)\pi_G(f2)\pi_G(d2) \\ \pi(g2) &= (0.8)(0.4)(0.1) + (0.85)(0.4)(0.9) + (0.3)(0.6)(0.1) + \\ &(0.25)(0.6)(0.9) \\ \pi(g2) &= 0.491\end{aligned}$$

Parte (b.1), percorrendo o segundo estado do nó G, que corresponde ao somatório das multiplicações dos valores da tabela de probabilidade pelas mensagens $\pi_X(z_i)$ que o filho G envia para seus pais F e D.

$$\begin{aligned}P(g1|\emptyset) &= \alpha\lambda(g1)\pi(g1) = \alpha(1)(0.509) = 0.797\alpha \\ P(g2|\emptyset) &= \alpha\lambda(g2)\pi(g2) = \alpha(1)(0.491) = 0.363\alpha\end{aligned}$$

Parte (b.2). Calcula a probabilidade de G quando nenhuma evidência é fornecida.

$$\begin{aligned}P(g1|\emptyset) &= \frac{0.509\alpha}{0.509\alpha + 0.491\alpha} = 0.509 \\ P(g2|\emptyset) &= \frac{0.491\alpha}{0.509\alpha + 0.491\alpha} = 0.491\end{aligned}$$

Parte (b) linha 1. Efetiva a normalização das probabilidades para que a

soma das mesmas seja 1. Estas são as probabilidades do nó G, após a inicialização da rede bayesiana.

As probabilidades dos nós F (Febre) e D (Dor de cabeça) na inicialização da rede bayesiana correspondem àquelas definidas em sua Tabela de Probabilidade Condicional (TPC), pois eles não possuem pais e os valores 0.509 e 0.491 são as probabilidades do nó G (Gripe) quando nenhuma evidência é fornecida. A rede bayesiana inicializada é mostrada na Figura 16, na qual são apresentados os valores λ e π de cada nó, bem como os valores das mensagens $\lambda_G()$ e $\pi_G()$.

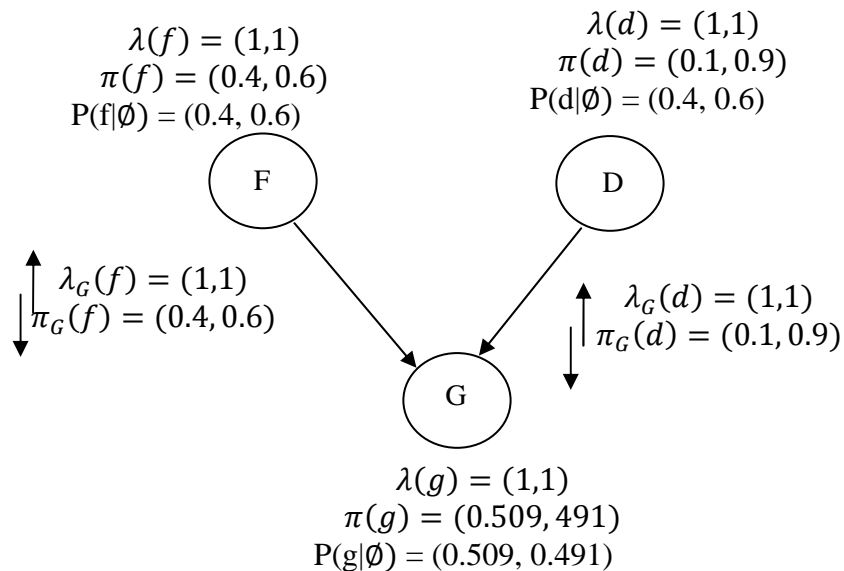


Figura 16. Rede Bayesiana Inicializada

Os cálculos a seguir consideram a propagação de probabilidades quando uma evidência é fornecida ao modelo. Utilizando a rede bayesiana descrita no exemplo anterior e considerando que o estado $g1$, tem gripe, do nó G, que representa gripe, é fornecido como evidência, as probabilidades atualizadas são obtidas utilizando o método *update-tree()*.

A chamada `update_tree(G, g1)` resulta nos seguintes passos:

$$A = \emptyset \cup \{G\} = \{G\};$$

$$A = \emptyset \cup \{g1\} = \{g1\};$$

Linha 1 do pseudocódigo. Esta linha adiciona o nó G ao conjunto de evidências e o estado g1 ao conjunto de estados.

$$\begin{aligned}\lambda(g1) &= 1; \pi(g1) = 1; P(g1|\{g1\}) = 1; \\ \lambda(g2) &= 0; \pi(g2) = 0; P(g2|\{g1\}) = 0;\end{aligned}$$

Linha 2. Quando uma evidência é fornecida, os valores associados ao respectivo estado são configurados para 1 e, executando o laço *para* da Linha 3, os valores para os demais estados são definidos como 0.

send_λ_msg(G, F);
send_λ_msg(G, D);

Execução do laço *para* situado na linha 5 do pseudocódigo.

A chamada *send_λ_msg(G, F)* resulta em:

$$\begin{aligned}\lambda_G(f1) &= [P(g1|f1, d1)\pi_G(d1) + P(g1|f1, d2)\pi_G(d2)]\lambda(g1) \\ &+ [P(g2|f1, d1)\pi_G(d1) + P(g2|f1, d2)\pi_G(d2)]\lambda(g2) \\ \lambda_G(f1) &= [(0.2)(0.1) + (0.15)(0.9)]1 + [(0.8)(0.1) + (0.85)(0.9)]0 \\ \lambda_G(f1) &= 0.155\end{aligned}$$

Cálculos correspondentes a parte (a.1). Nesta etapa, é calculada a mensagem que o nó filho G envia para seu pai F. Primeiramente, esta equação calcula o somatório das multiplicações entre a tabela de probabilidades do nó e as mensagens $\pi_G(D)$ que o pai D envia para seu filho G. Logo após, é realizado o somatório das multiplicações entre os valores resultantes do somatório anterior e os valores $\lambda(G)$. Considera-se, no lado direito da equação, somente os pais de G distintos de F, neste exemplo, o nó D.

$$\begin{aligned}\lambda_G(f2) &= [P(g1|f2, d1)\pi_G(d1) + P(g1|f2, d2)\pi_G(d2)]\lambda(g1) \\ &+ [P(g2|f2, d1)\pi_G(d1) + P(g2|f2, d2)\pi_G(d2)]\lambda(g2) \\ \lambda_G(f2) &= [(0.7)(0.1) + (0.75)(0.9)]1 + [(0.3)(0.1) + (0.25)(0.9)]0 \\ \lambda_G(f2) &= 0.745\end{aligned}$$

Realização dos cálculos da parte (a.1) para o estado f2 do nó F.

$$\lambda(f1) = \lambda_G(f1) = 0.155;$$

$$\lambda(f2) = \lambda_G(f2) = 0.745;$$

Percorrendo todos os estados do nó F e executando a parte (a.2) do pseudocódigo, obtêm-se os novos valores de $\lambda(F)$.

$$P(f1|\{g1\}) = \alpha\lambda(f1)\pi(f1) = \alpha(0.155)(0.4) = 0.062\alpha$$

$$P(f2|\{g1\}) = \alpha\lambda(f2)\pi(f2) = \alpha(0.745)(0.6) = 0.447\alpha$$

Parte (a.3). Cálculo das probabilidades de F quando a evidência $g1$ é fornecida.

$$P(f1|\{g1\}) = \frac{0.062\alpha}{0.062\alpha + 0.447\alpha} = 0.1218$$

$$P(f2|\{g1\}) = \frac{0.447\alpha}{0.062\alpha + 0.447\alpha} = 0.8781$$

Parte (b), linha 1. Normalização dos resultados e atribuição dos valores à tabela de probabilidade condicional do nó F, ou seja, a probabilidade de ter febre dado que tem gripe.

A chamada *send_λ_msg(G, D)* resulta em:

$$\lambda_G(d1) = [P(g1|f1, d1)\pi_G(f1) + P(g1|f2, d1)\pi_G(f2)]\lambda(g1) \\ + [P(g2|f1, d1)\pi_G(f1) + P(g2|f2, d1)\pi_G(f2)]\lambda(g2)$$

$$\lambda_G(d1) = [(0.2)(0.4) + (0.7)(0.6)]1 + [(0.8)(0.4) + (0.3)(0.6)]0 = 0.5;$$

Cálculos correspondentes a parte (a.1). Nesta etapa, é calculada a mensagem que o nó filho G envia para seu pai D. Conforme anteriormente descrito, considera-se no lado direito da equação somente os pais de G distintos de D, neste exemplo, o nó F, conforme ilustra esta parte do pseudocódigo.

$$\lambda_G(d2) = [P(g1|f1, d2)\pi_G(f1) + P(g1|f2, d2)\pi_G(f2)]\lambda(g1) \\ + [P(g2|f1, d2)\pi_G(f1) + P(g2|f2, d2)\pi_G(f2)]\lambda(g2)$$

$$\lambda_G(d2) = [(0.15)(0.4) + (0.75)(0.6)]1 + [(0.85)(0.4) + (0.25)(0.6)]0 \\ \lambda_G(d2) = 0.51$$

Realização dos cálculos da parte (a.1) para o estado $d2$ do nó D.

$$\lambda(d1) = \lambda_G(d1) = 0.5;$$

$$\lambda(d2) = \lambda_G(d2) = 0.51;$$

Percorrendo todos os estados do nó D e executando a parte (a.2) do pseudocódigo, obtêm-se os novos valores de $\lambda(D)$.

$$\begin{aligned} P(d1|\{g1\}) &= \alpha\lambda(d1)\pi(d1) = \alpha(0.5)(0.1) = 0.05\alpha \\ P(d2|\{g1\}) &= \alpha\lambda(d2)\pi(d2) = \alpha(0.51)(0.9) = 0.459\alpha \end{aligned}$$

Parte (a.3). Cálculo das probabilidades de D quando a evidência $g1$ é fornecida, ou seja, a probabilidade de ter dor de cabeça dado que tem gripe.

$$\begin{aligned} P(d1|\{g1\}) &= \frac{0.05\alpha}{0.05\alpha + 0.459\alpha} = 0.09823 \\ P(d2|\{g1\}) &= \frac{0.459\alpha}{0.05\alpha + 0.459\alpha} = 0.90176 \end{aligned}$$

Parte (b), linha 1. Normalização dos resultados e atribuição dos valores à tabela de probabilidade condicional do nó D.

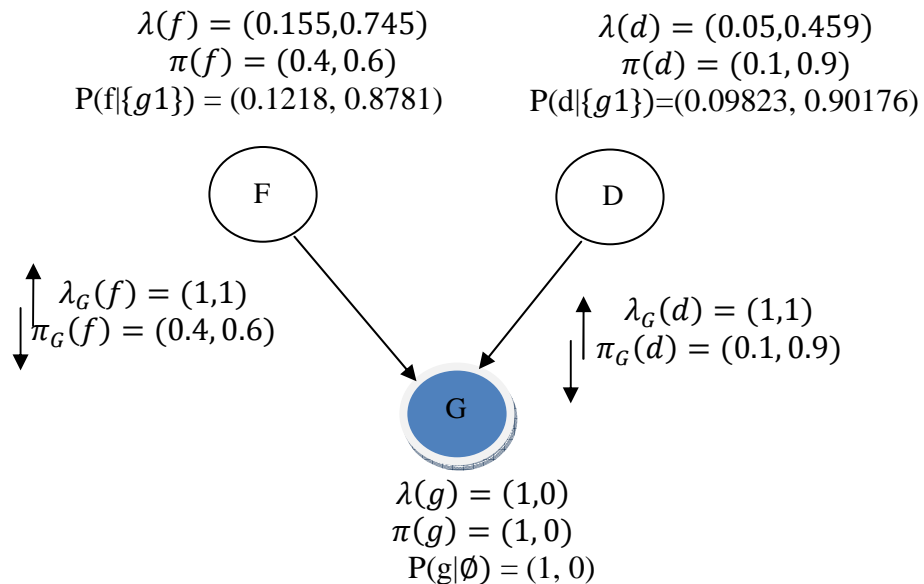


Figura 17. Probabilidades da rede bayesiana após a evidência $g1$

A Figura 17 apresenta as probabilidades da rede bayesiana após o fornecimento da evidência $g1$, ou seja, está com gripe. As probabilidades $P(f|\{g1\}) = (0.1218, 0.8781)$ e $P(d|\{g1\}) = (0.09823, 0.90176)$ são as probabilidades

condicionais dos nós F(tem febre e não tem febre) e D (tem dor de cabeça e não tem dor de cabeça) dado a ocorrência do estado tem febre (g1) do nó G (gripe).

7.3 MODELAGEM DA SHELL

Além da modelagem e desenvolvimento de uma *shell* para redes bayesianas, esta pesquisa compreende a diagramação em *Unified Modeling Language* (UML)¹⁸ da *shell* Pegasus, na qual os resultados deste trabalho serão posteriormente integrados.

Foram desenvolvidos diagramas tanto relacionados a *shell* Pegasus (casos de uso e atividade) quanto ao módulo de redes bayesianas (diagramas de classe, de caso de uso e atividades), os quais foram criados utilizando um *plugin* UML para o ambiente de desenvolvimento Netbeans 6.1, gratuitamente disponível em <http://www.netbeans.org>.

Primeiramente são apresentados os diagramas relacionados à *shell* Pegasus. Conforme ilustra a Figura 18, a utilização básica desta *shell* consiste na escolha da teoria utilizada para resolver o problema (fatores de certeza, *fuzzy*, dempster-shafer e redes bayesianas), e o algoritmo pertencente a esta teoria responsável pela execução do processo de inferência. Em seguida, os parâmetros de entrada são fornecidos e a solicitação para realizar uma inferência é executada. Assim, o mecanismo de inferência correspondente à teoria selecionada é acionado e executado, sendo seus resultados processados e encaminhados para visualização.

¹⁸ Notação para modelagem de sistemas, utilizando conceitos orientados a objetos (LARMAN, 2000)

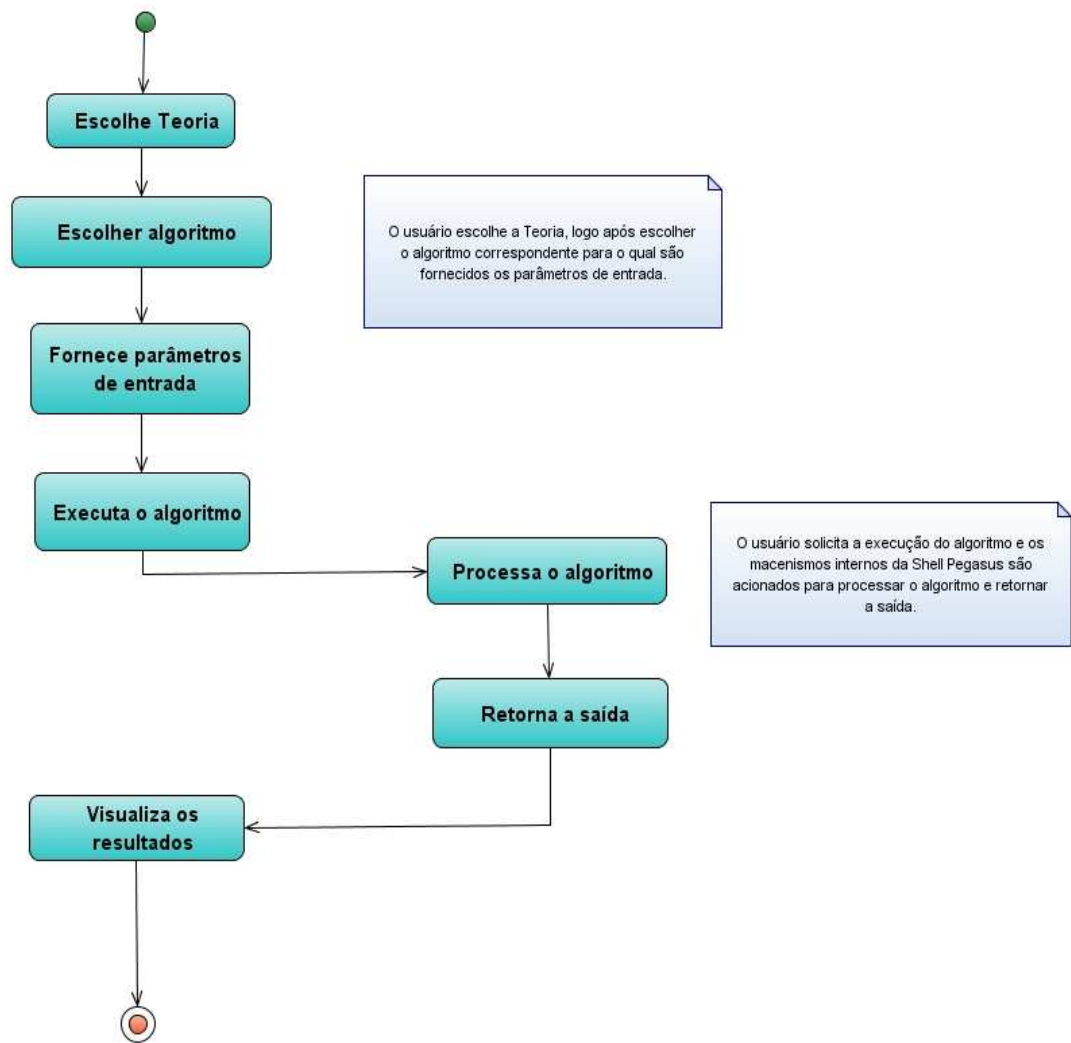


Figura 18. Diagrama de atividades da *shell* Pegasus

A Figura 19 apresenta um diagrama de casos de uso da ferramenta, onde o usuário pode escolher entre as diferentes teorias disponibilizadas, sendo a *shell* Pegasus responsável por acionar a teoria correspondente, ação esta que compreende a abertura da interface correspondente ao mecanismo de modelagem da incerteza selecionado e a conseqüente disponibilização dos recursos fornecidos pela mesma.

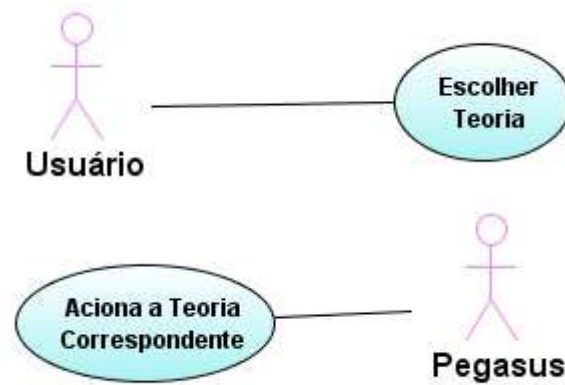


Figura 19. Diagrama de casos de uso da *shell* Pegasus

Os diagramas UML apresentados descrevem como utilizar a *shell* Pegasus, assim, a seguir são apresentados e descritos os correspondentes do módulo de redes bayesianas desta ferramenta. A Figura 20 ilustra o diagrama de classes associado à modelagem da implementação das redes bayesianas.

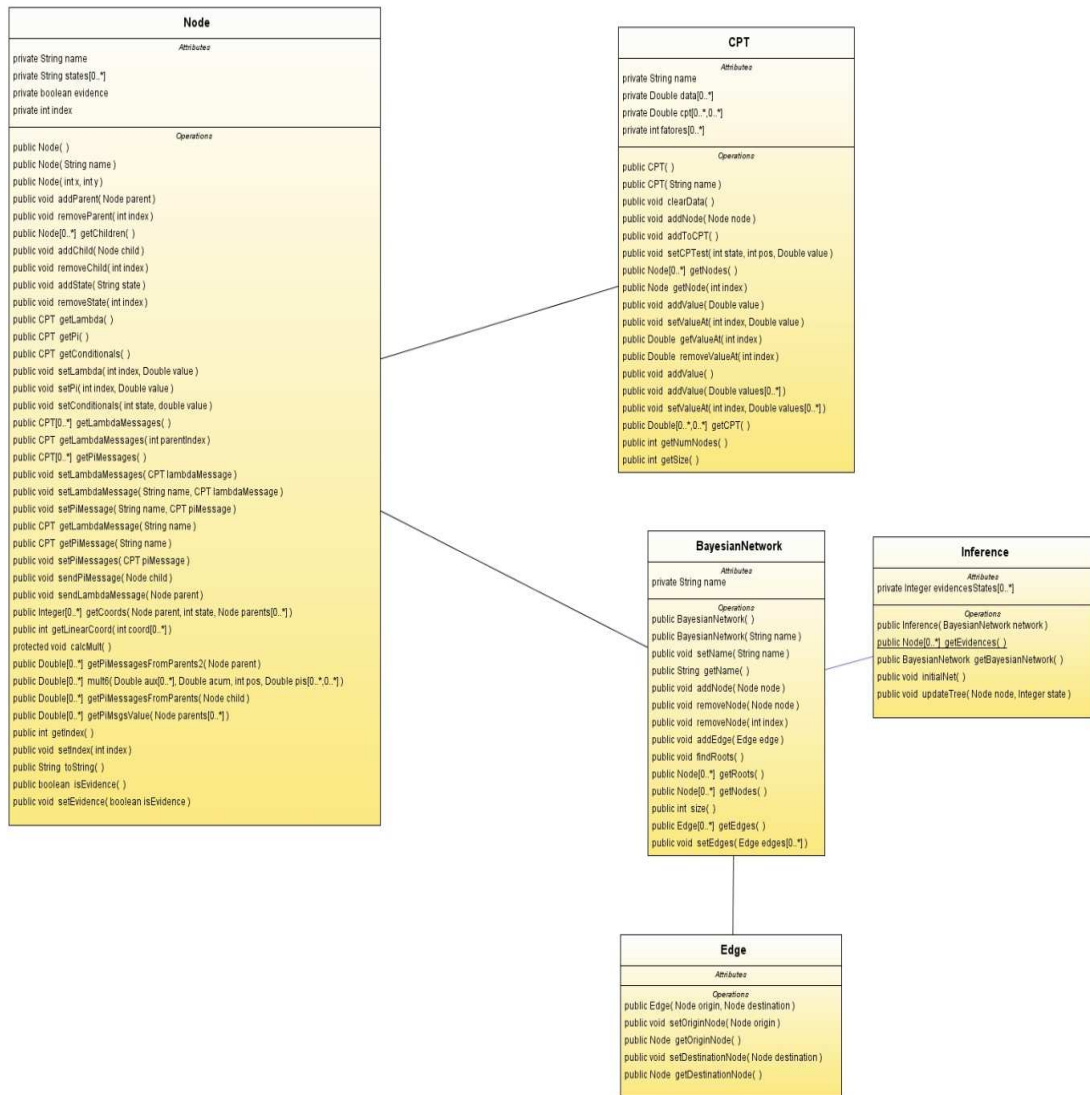


Figura 20. Diagrama de classes do módulo bayesiano

Neste diagrama constam as classes fundamentais para efetivação de inferências em determinado modelo probabilístico utilizando o algoritmo de passagem de mensagens de Pearl (PEARL, 1988). A classe *Node* é responsável pela implementação dos métodos *send_pi_msg()*, *send_lambda_msg()* e os métodos necessários para a implementação dos mesmos. A classe *CPT*¹⁹ abstrai uma tabela de probabilidade a *priori* e condicional, sendo que a cada nó é associado uma *CPT* correspondente.

A classe *Edge* é utilizada para armazenar o padrão de conectividade da rede

¹⁹ Conditional Probability Table corresponde a Tabela de Probabilidade Condicional do nó (TPC).

bayesiana, identificando os nós origem e destino de cada arco inserido no modelo. Já a classe *Inference* é responsável pelo desenvolvimento dos métodos *initial_net()* e *update_tree()* e finalizando esta descrição, a classe *BayesianNetwork* representa as partes qualitativas e quantitativas de uma rede bayesiana. A construção de um modelo bayesiano utilizando esta ferramenta é descrita pelo diagramas de casos de uso e atividades.

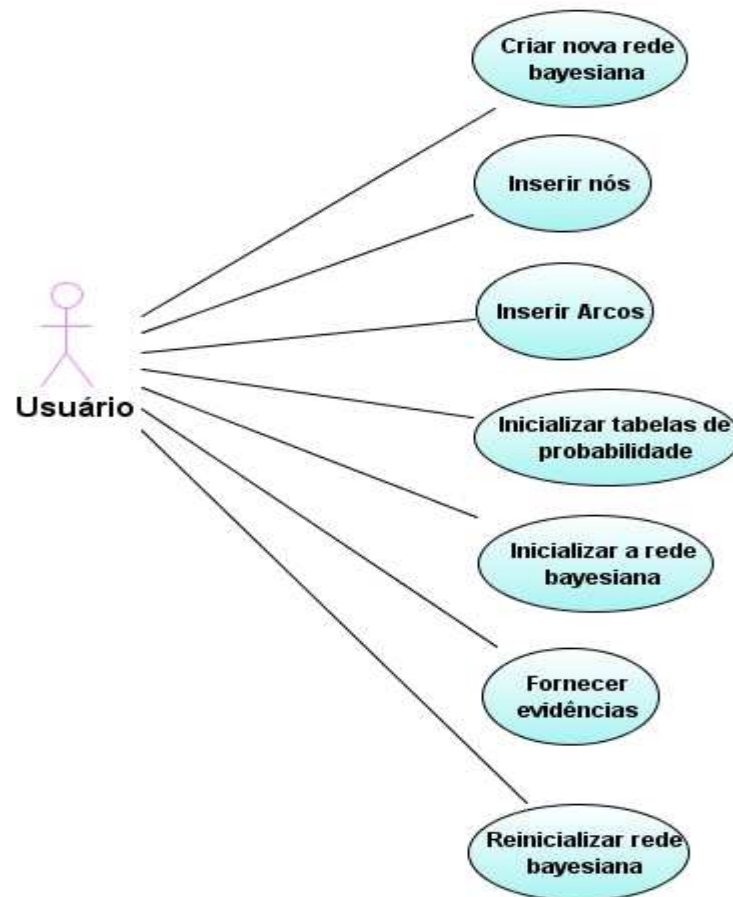


Figura 21. Diagrama de casos de uso da ferramenta de redes bayesianas

O diagrama ilustrado na Figura 21 apresenta as possibilidades de ação do usuário no módulo bayesiano, que convergem para a criação de uma rede bayesiana por meio da inserção de nós, que representam as variáveis, de arcos que caracterizam o relacionamento causal entre os mesmos, definindo desta forma o componente

qualitativo do modelo sendo criado. Em seguida, ao usuário fornece os valores das tabelas de probabilidade *a priori* e condicional, especificando, portanto, o componente quantitativo, possibilitando a inicialização da rede e o posterior fornecimento de evidências. Quando necessário, o usuário pode reinicializar a distribuição de probabilidades da rede bayesiana.

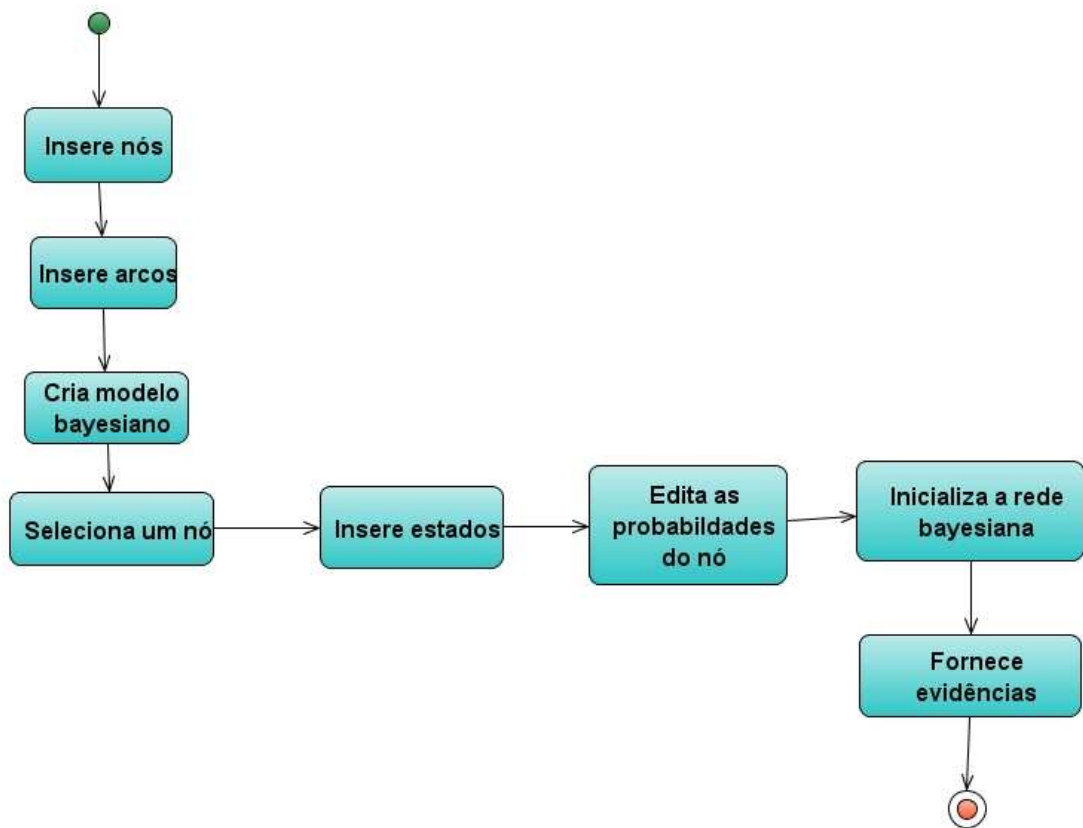


Figura 22. Diagrama de atividades da ferramenta de redes bayesianas

Conforme ilustra a Figura 22, o usuário inicialmente cria os nós que constituem o modelo que se está desenvolvendo, logo após os arcos e, assim a estrutura da rede é definida. Os estados do nó selecionado são então inseridos juntamente com as probabilidades de sua CPT. Após estes procedimentos, a rede bayesiana pode ser inicializada e atualizada à medida que novas evidências são fornecidas.

7.4 IMPLEMENTAÇÃO DO MÓDULO BAYESIANO E DA SHELL PEGASUS

A *shell* Pegasus é uma ferramenta baseada em conceitos de inteligência artificial cujo objetivo é auxiliar na modelagem de problemas que empregam o conceito de incerteza. O desenvolvimento das teorias anteriormente citadas foi realizado utilizando a linguagem de programação Java, devido suas características multiplataforma e *multithreading* bem como bons ambientes de desenvolvimento, como o NetBeans 6.1, escolhido para esta pesquisa.

Utilizando convenções de codificação da linguagem da Java²⁰, como nomes de atributos, métodos, a aplicação foi projetada para execução em um ambiente desktop e devido a características próprias de Java, multiplataforma.

Considerando especificamente o módulo de redes bayesianas, sua implementação é baseada no pseudocódigo descrito na seção 7.2 (NEOPOLITAN, 2004). Este algoritmo propicia o desenvolvimento de uma metodologia de propagação do teorema de Bayes em uma rede bayesiana com estrutura de poliárvore por meio da exploração de dependências e independências condicionais que uma vez desenvolvido, precisa ser testado e seu funcionamento devidamente averiguado para que possa ser efetivamente utilizado.

7.5 APLICAÇÃO DO MÓDULO BAYESIANO A UMA PROBLEMÁTICA DA ÁREA DA SAÚDE

Os testes deste módulo da *shell* Pegasus foram realizados utilizando a questão de diagnóstico de lombalgia.

Existem cerca de cinquenta doenças relacionadas à região lombar, das quais

²⁰ <http://java.sun.com/docs/codeconv/>

seis serão diagnosticadas, devido sua alta prevalência: espondilite ancilósante, artrite (osteoartrite ou osteoartrose), artrite reumatóide, espondilolistese, hérnia discal e lombalgia por causa muscular. Esta base foi escolhida devido sua utilização em outro trabalho e, portanto, possibilitando a comparação dos resultados. Para obter maiores detalhes sobre a lombalgia, esta pesquisa possui um anexo correspondente ao assunto retirado do trabalho de Telma Scussel, da Universidade do Extremo Sul Catarinense (UNESC), desenvolvido no ano de 2001. Maiores informações vide ANEXO A ou o trabalho (SCUSSEL, 2001).

Uma vez definidas as doenças a serem diagnosticadas, os testes deste módulo da *shell* Pegasus foram realizados utilizando uma base de conhecimentos para o diagnóstico de lombalgia, desenvolvida em 2001 (SCUSSEL, 2001), a qual foi criada utilizando a *shell* Netica tendo-se como especialista o médico Robson Luiz dos Santos, professor do curso de Medicina da Universidade do Extremo Sul Catarinense (UNESC). Os sinais e sintomas encontrados nesta base, a partir da qual o modelo probabilístico será desenvolvido são ilustrados nas tabelas a seguir. A Tabela 5 apresenta variáveis relacionadas ao paciente como sexo, idade e histórico familiar.

Tabela 5. Variáveis relacionadas ao paciente

Sexo	Masculino
	Feminino
Idade	20-30
	31-40
	41-50
	51-60
	61-70
Histórico familiar	Apresenta patologia
	Não apresenta patologia

Fonte: SCUSSEL, T. (2001)

Em relação aos hábitos de vida, a tabela 6 apresenta a prática de esporte, o

trabalho pesado e a presença de trauma.

Tabela 6. Variáveis dispostas segundo os hábitos de vida

Prática de esporte	Sim
	Não
Trabalho pesado	Sim
	Não
Presença de trauma	Sim
	Não

Fonte: SCUSSEL, T. (2001)

Em relação à dor, têm-se as variáveis início, caráter da dor, irradiação, parestesia da dor, déficit motor, sensitiva e clima:

Tabela 7. Variáveis organizadas pelo critério dor

Início	Aguda
	Incidiosa
Caráter dor	Profundo
	Visceral
	Superficial
Irradiação	Sim
	Não
Parestesia da dor	Sim
	Não
Déficit motor	Sim
	Não
Sensitiva	Sim
	Não
Clima	Sim
	Não

Fonte: SCUSSEL, T. (2001)

Em relação ao exame físico, têm-se as variáveis palpação, dor para vertebral, trajeto ciático e contratura muscular.

Tabela 8. Variáveis organizadas por exame físico

Palpação	Sim
	Não
Dor para vertebral	Sim
	Não

Trajetos ciáticos	Sim
	Não
Contratura muscular	Sim
	Não

Fonte: SCUSSEL, T. (2001)

Em relação ao exame neurológico têm-se como variáveis relacionadas a diminuição de sensibilidade nos membros inferiores, diminuição de força e reflexos.

Tabela 9. Variáveis agrupadas segundo o exame neurológico

Diminuição de sensibilidade nos membros inferiores	Sim
	Não
Diminuição de força	Sim
	Não
Reflexos	Sim
	Não

Fonte: SCUSSEL, T. (2001)

7.6 RESULTADOS OBTIDOS

O resultado desta pesquisa é uma ferramenta para construção de redes bayesianas e efetivação de inferências utilizando o algoritmo de poliárvores de Judea Pearl desenvolvido em 1988 (PEARL, 1988). A construção de um modelo probabilístico é realizada visualmente por meio da especificação dos nós e arcos que definem o relacionamento causal inerente ao modelo. A Figura 22 apresenta a interface gráfica inicial da ferramenta.

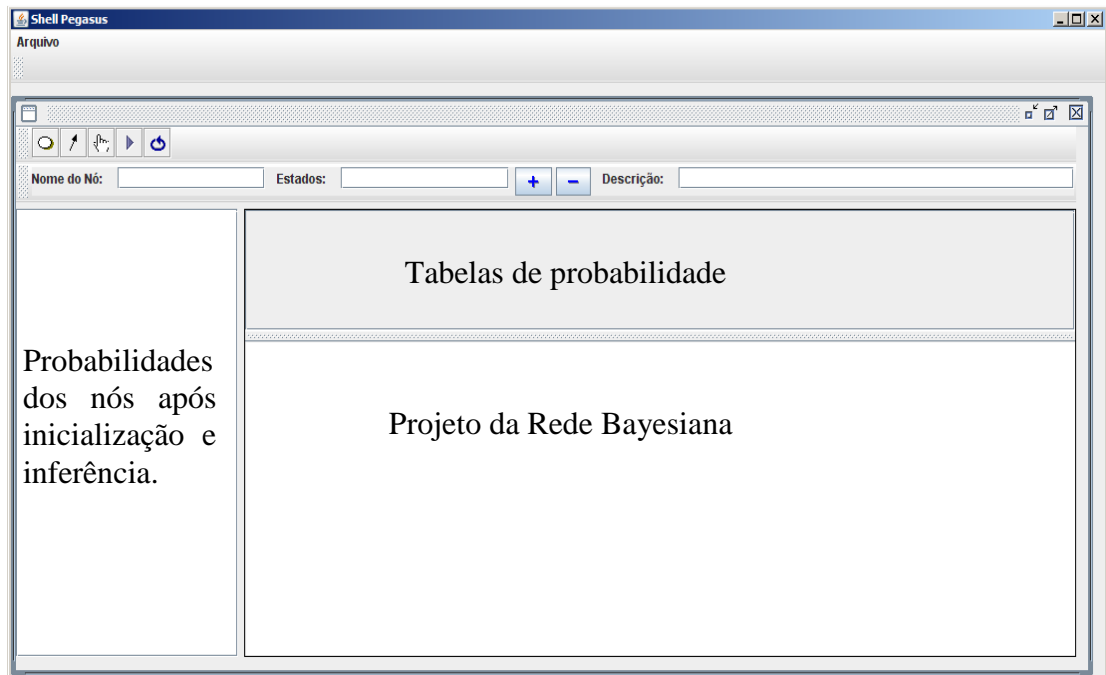


Figura 23. Interface inicial da ferramenta

A interface consiste basicamente em três regiões principais. Uma onde a rede bayesiana é desenhada, outra onde as tabelas de probabilidades *a priori* ou condicionais são criadas e inicializadas, e a terceira região apresenta as probabilidades do modelo depois de inicializado ou após o fornecimento de alguma evidência.

A Figura 24 apresenta a barra de ferramentas utilizada para projetar as redes bayesianas utilizando a *shell*.

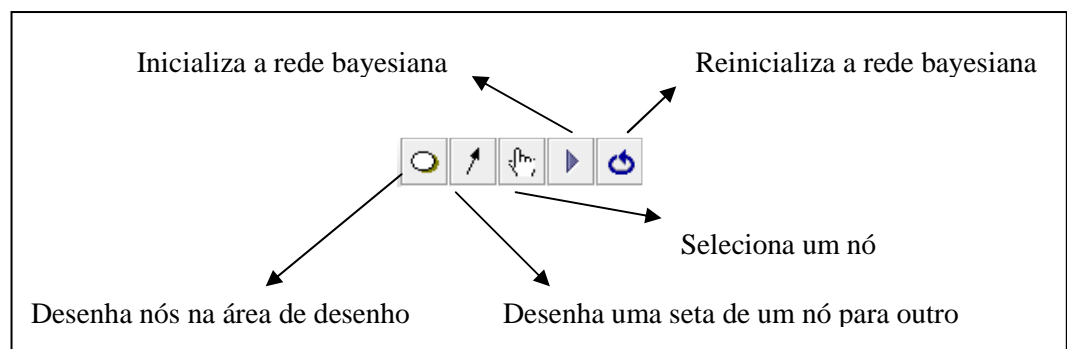


Figura 24. Barra de ferramentas para trabalhar com redes bayesianas

A Figura 25 ilustra o meio pelo qual determinados atributos do nó, como nome, estados e descrição são definidos.



Figura 25. Barra para inicializar os atributos do nó

Após a definição das variáveis e seus respectivos estados, a rede bayesiana para o diagnóstico de lombalgia pode ser representada utilizando o módulo bayesiano da *shell* Pegasus. A Figura 25 ilustra este modelo, com o nó S19, correspondente a variável idade, selecionado.

Os nós são identificados por nomes automaticamente atribuídos, conforme apresentado em outras ferramentas, com o objetivo de facilitar a criação da rede bayesiana, porém, conforme a necessidade do usuário é possível alterá-lo tendo como consequência, um efeito visual não padronizado, pois nesta versão da *shell* o diâmetro da circunferência que representa o nó não é ajustável em função de seu respectivo nome.

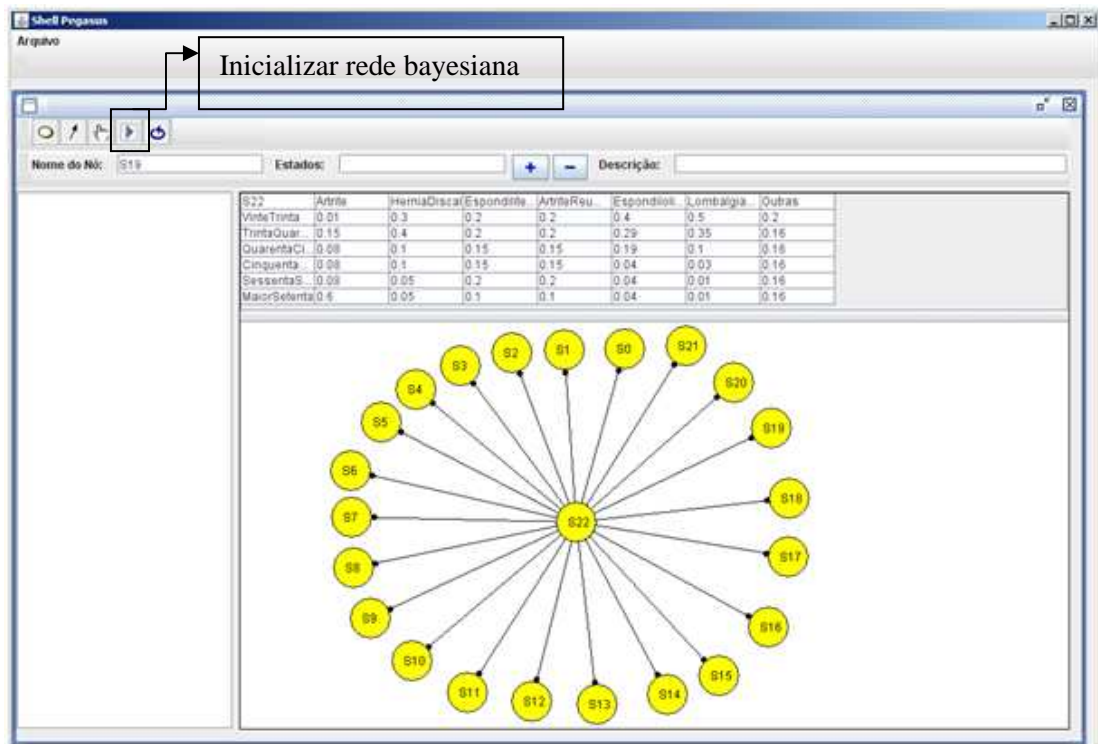



Figura 26. Rede bayesiana para o diagnóstico de lombalgia.

Cada nó, com exceção dos nós raiz que possuem uma tabela de

probabilidades *a priori*, possui uma tabela de probabilidades condicionais que deve ser inicializada com as respectivas probabilidades antes que quaisquer processos de inicialização e consulta no modelo sejam executados.

No momento em que todas as tabelas foram especificadas, é necessário clicar no botão  para inicializar as probabilidades da rede bayesiana. A Figura 26 ilustra a rede bayesiana após a propagação das probabilidades por sua estrutura. Nesta figura, o nó Lombalgia (S22), está selecionado. É neste nó que constam as doenças localizadas na região lombar a serem diagnosticadas, cuja distribuição é atualizada quando uma evidência é fornecida. É importante observar os estados do nó Lombalgia (S22), ilustrado na Figura 27, também podem ser fornecidos como evidência para obter as probabilidades dos estados dos demais nós que compõe o modelo.

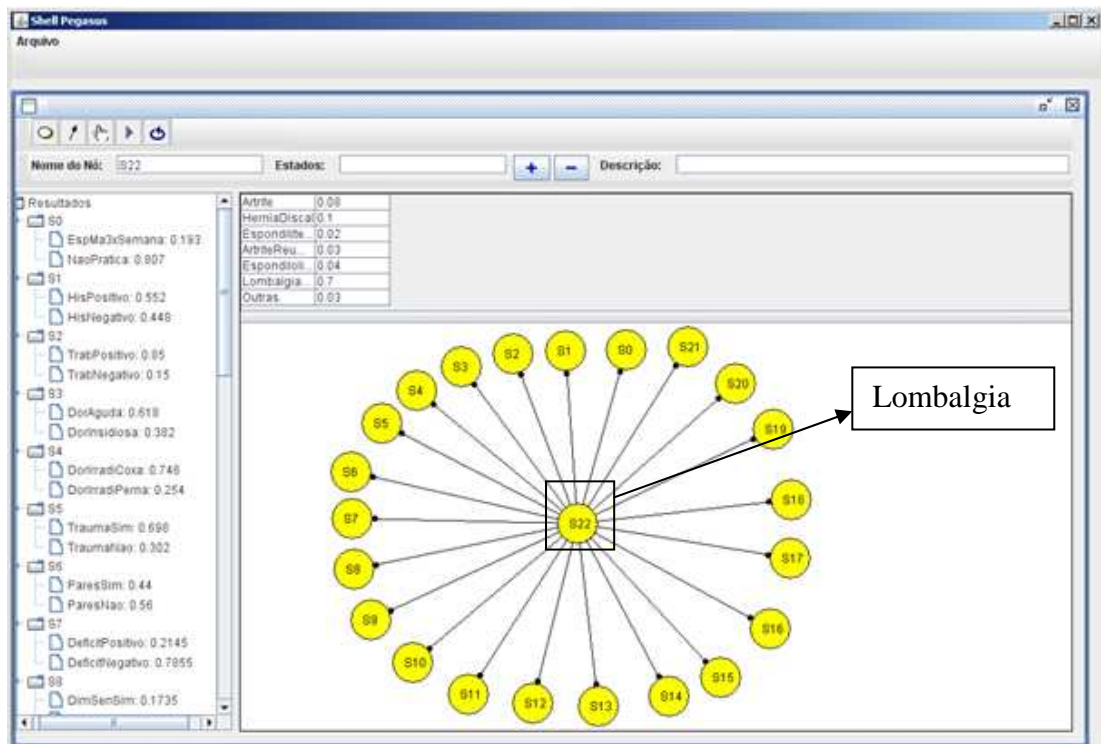


Figura 27. Rede bayesiana inicializada

A tabela de probabilidades do nó Lombalgia (S22) quando nenhuma evidência é fornecida é ilustrada na Figura 28.

Artrite	0.08
HerniaDiscal	0.1
Espondilite...	0.02
ArtriteReu...	0.03
Espondiloli...	0.04
Lombalgia ...	0.7
Outras	0.3

Figura 28. Probabilidades quando nenhuma evidência é fornecida

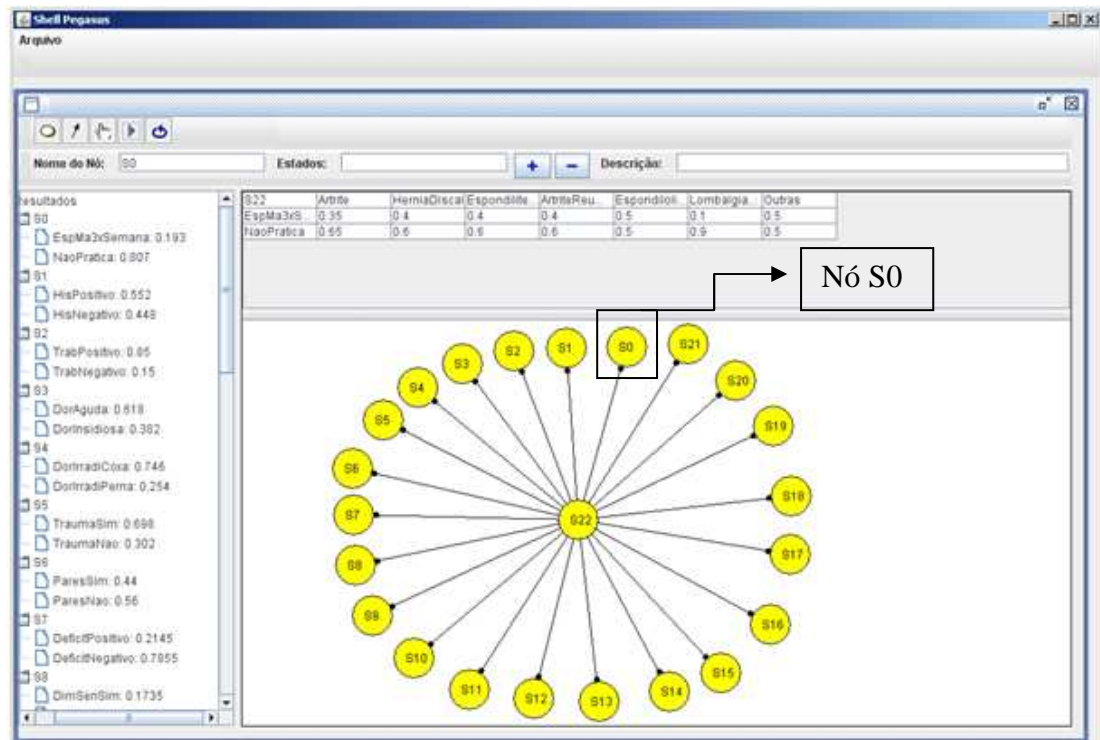


Figura 29. Modelo bayesiano desenvolvido com o nó S0 selecionado

Uma evidência é fornecida clicando duas vezes no estado da respectiva variável. Automaticamente, o estado selecionado assume o valor 1 enquanto os demais são configurados para zero e a distribuição de probabilidades do modelo é atualizada refletindo a ocorrência da evidência. A Figura 30 apresenta as probabilidades do nó Lombalgia (S22) quando a evidência Histórico Familiar (S1) assume o estado Positivo. A Figura 31 ilustra, na área da interface de exibição de probabilidades, o fornecimento desta evidência, alterando o valor do estado positivo para 1.

Artrite	0.0942
HerniaDiscal	0.1178
Espondilite...	0.0254
ArtriteReu...	0.038
Espondiloli...	0.0507
Lombalgia ...	0.6467
Outras	0.0272

Figura 30. Probabilidades quando histórico familiar positivo

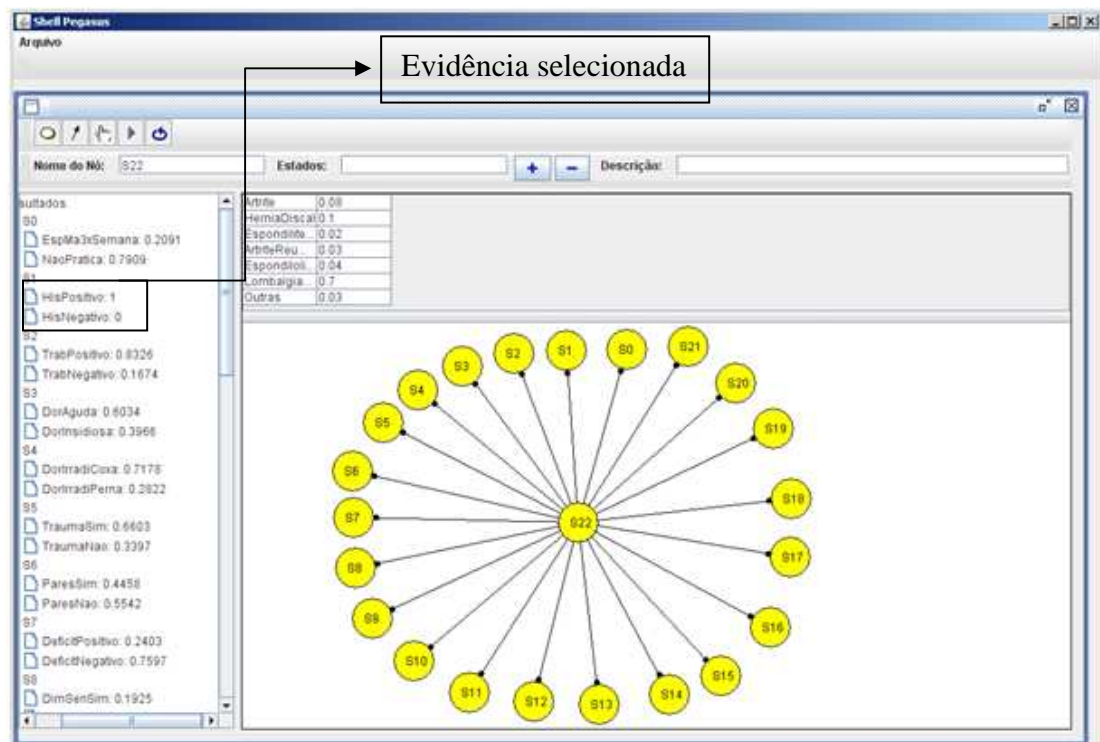


Figura 31. Rede bayesiana quando histórico familiar positivo

Em seguida, fornecendo a evidência Dor no estado Aguda, as probabilidades da rede bayesiana são atualizadas, tanto para a evidência Histórico Familiar assumindo o estado Positivo, quanto para Dor em seu estado Aguda, as quais são ilustradas na Figura 32.

Artrite	0.0156
HerniaDiscal	0.1171
Espondilite...	0.0126
ArtriteReu...	0.0441
Espondiloli...	0.0378
Lombalgia ...	0.7502
Outras	0.0225

Figura 32. Probabilidades quando histórico familiar positivo e dor aguda

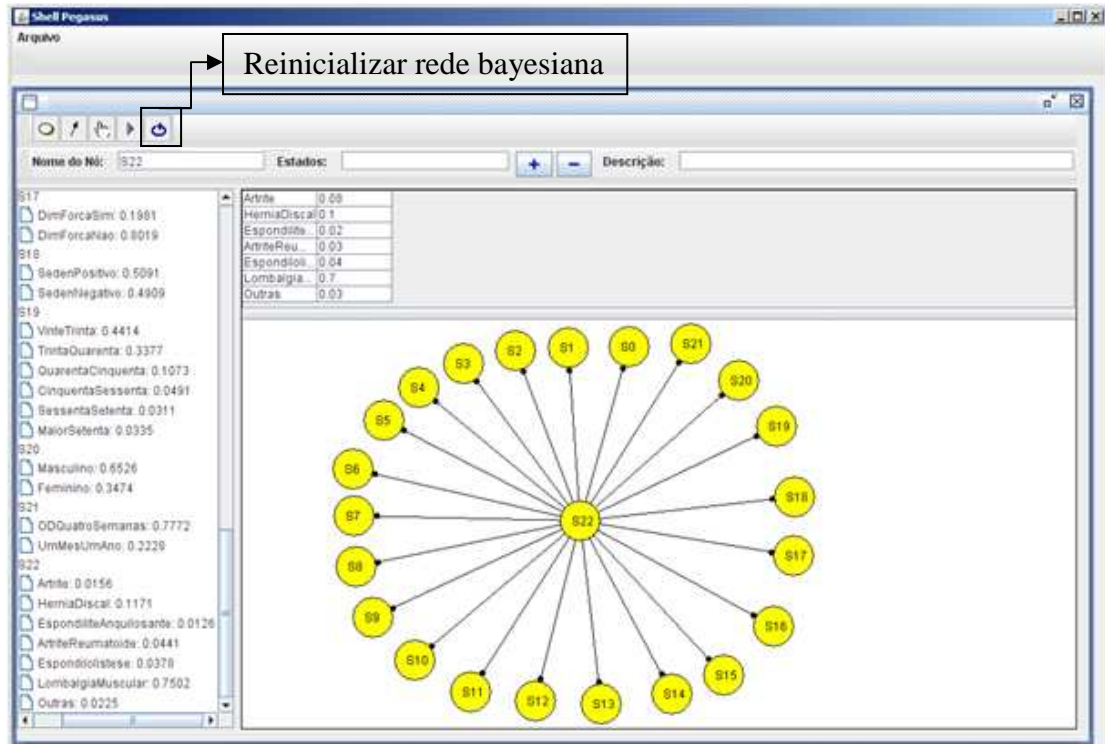



Figura 33. Modelo quando histórico familiar positivo e dor aguda

É possível fornecer tantas evidências quantas forem as disponíveis e avaliar os resultados das consultas realizadas, além da possibilidade de removê-las quando necessário, função realizada pelo botão , cuja atribuição é reinicializar a rede bayesiana ao seu estado original, ou seja, quando nenhuma evidência é fornecida.

Os resultados obtidos pelo módulo bayesiano desenvolvido são condizentes com os apresentados pela *shell* Netica, ferramenta na qual a base de conhecimento de lombalgia foi desenvolvida (SCUSSEL, 2001), o que indica que o mecanismo de propagação de probabilidades implementado apresenta correto funcionamento para efetivação de inferências em redes bayesianas com estruturas de poliárvore.

No entanto, devido ao algoritmo desenvolvido ser projetado para poliárvores, as redes bayesianas nas quais ele é aplicado devem obedecer a esta estrutura, caso contrário, os resultados obtidos serão incorretos. Uma vez explicitada a limitação algorítmica relacionada aos modelos construídos, também são descritas

algumas limitações de interface que ainda não foram desenvolvidas, tais como a impossibilidade de arrastar os nós e arcos da rede, a direcionalidade dos arcos ser representada por elipses ao invés de setas, o não redimensionamento automático dos nós em função do número de caracteres de seu nome e a adequação da largura das células das tabelas de probabilidade ao seu conteúdo.

Este capítulo descreveu, de forma detalhada, a pesquisa desenvolvida. Foram apresentadas as etapas metodológicas para sua realização bem como a apresentação dos resultados obtidos. O próximo capítulo conclui esta pesquisa e propõe direcionamentos para trabalhos futuros.

8 CONCLUSÃO

A modelagem da incerteza caracteriza um mecanismo para solucionar problemas nos quais suas diferentes manifestações se fazem presentes. Quando o relacionamento existente entre as variáveis do modelo que está sendo desenvolvido remete naturalmente a um processo de raciocínio probabilístico, como, por exemplo, a probabilidade de gripe sendo que apresenta o sintoma febre, as redes bayesianas se tornam uma boa alternativa e devido a esta característica, é amplamente utilizada.

Este é apenas um exemplo no qual a incerteza por aleatoriedade está presente e precisa ser adequadamente tratada. Assim, existem inúmeras aplicações do formalismo bayesiano, especialmente na área da saúde, sendo que o enfoque principal desta pesquisa foi o desenvolvimento de uma ferramenta para redes bayesianas que disponibilizou um mecanismo de inferência para poliárvores a partir do algoritmo de passagem de mensagens de Pearl (PEARL, 1988).

A utilização de um exemplo numérico buscou auxiliar na compreensão das etapas envolvidas na aplicação do algoritmo e, ao final deste trabalho, a comunidade de inteligência artificial será munida com uma nova ferramenta para desenvolvimento de sistemas especialistas probabilísticos, os quais poderão ser desenvolvidos tanto no ambiente do módulo bayesiano da *shell* Pegasus como em aplicações distintas utilizando a API disponibilizada.

Considerando os resultados obtidos na avaliação da ferramenta, a qual foi aplicada ao problema de diagnóstico de lombalgia, os resultados mostraram-se condizentes com aqueles apresentados anteriormente (SCUSSEL, 2001) e, portanto, satisfatórios, segundo a opinião do especialista, o que motiva prosseguir no desenvolvimento desta ferramenta.

Neste sentido, diversas são as alternativas para desenvolvimento de trabalhos futuros na temática desta pesquisa, tanto em mecanismos de inferência, aprendizado e redes de decisão. Dentre os algoritmos de inferência que poderiam ser desenvolvidos encontra-se o Junction Tree (JENSEN, 2001), utilizado na *shell* Netica e no Hugin Expert, para o aprendizado, diversos algoritmos tanto baseados no método de busca e pontuação (K2) como os baseados em independência condicional (PC) e a disponibilização de redes de decisão, seriam direcionamentos para trabalhos futuros, além da validação da base de conhecimento de lombalgia e um projeto de interface levando em consideração características de usabilidade e ergonomia.

REFERÊNCIAS

- ALMEIDA, P. E.; EVSUKOFF, A. G. Sistemas Fuzzy. In: REZENDE, S.O.(Org.) **Sistemas Inteligentes: fundamentos e aplicações**. Barrueri: Manole, 2005. p. 169-201.
- BADIRU, Adedeji Bodunde; CHEUNG, John Y. **Fuzzy engineering expert systems with neural network applications**. New York: J. Wiley, 2002.
- BARBETTA, Pedro Alberto; REIS, Marcelo Menezes; BORNIA, Antonio Cezar. **Estatística: para cursos de engenharia e informática**. São Paulo: Atlas, 2004. 410 p
- BARRETO, J. M. **Inteligência artificial no limiar do século XXI**. 3.ed. Florianópolis: Duplic, 2001.
- BAYESIA - <http://www.bayesia.com/Docs/gbTutorialBLab.pdf>
- BITTENCOURT, Guilherme. **Inteligência artificial: ferramentas e teorias**. Florianópolis: UFSC, 2004.
- BRAGA, Antônio de Pádua; CARVALHO, André Carlos Ponce de Leon Ferreira de; LUDERMIR, Teresa Bernarda. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2000.
- CASTILHO, Enrique; GUTIÉRREZ, José Manuel; HADI, Ali S. **Sistemas Expertos y Modelos de Redes Probabilísticas**. Madrid: Academia Española de Ingeniería, 1998.
- COSTA, E.; SIMÕES, A. **Inteligência Artificial: fundamentos e aplicações**. Lisboa: FCA, 2004.
- COSTA, Paulo Afonso Bracarense. **Um enfoque segundo a teoria dos conjuntos difusos para a meta-análise**. 1999. 140 f. Tese (Doutorado) - Ufsc, Florianópolis, 2001.
- COX, Earl. **Fuzzy modeling and genetic algorithms for data mining and exploration**. California: Morgan Kaufmann, 2005.
- da SILVA, W. T.; LADEIRA, M. **Mineração de dados em Redes Bayesianas**. Organizado por: Ingrid Jansch Porto Jornada de Atualização em Informática - XXI JAI; in anais do XXII Congresso Brasileiro de Computação., Florianópolis: SBC, 2002., Editado por Ingrid Jansch Porto, 2002, v. 2, p. 235-286

DAVIS, R.; KING, J. J. The Context of the MYCIN Experiments. In: BUCHANAN, B. G.; SHORTLIFFE, E. H. (Org). **Rule-Based Expert Systems: the MYCIN Experiments of the Stanford Heuristic Programming Project**. California: Addison-Wesley, 1984. p. 3-19.

DEITEL, H.M; DEITEL, P. J. **Java: como programar**. 6.ed São Paulo: Pearson Education do Brasil, 2005.

DIEZ, Francisco Javier. **Introducción al Razonamiento Aproximado**. Disponível em: <<http://www.ia.uned.es/%7Efjdiez/libros/razaprox.pdf>>. Acesso em: 2 set. 2007.

FERRARI, Giselle Lopes. **Intellec: shell para desenvolvimento de sistemas especialistas**. 2005. 87 f. Dissertação (Mestrado) - Ufsc, Florianópolis, 2005.

JENSEN, Finn V.. **Bayesian networks and decision graphs**. New York: Springer, 2001.

KLIR, George; YUAN, Bo. **Fuzzy sets and fuzzy logic: theory and applications**. New Jersey: Prentice Hall, 1995.

KORB, Kevin; NICHOLSON, Ann. **Bayesian artificial intelligence**. Florida: Chapman & Hall, 2004.

LARMAN, Graig. **Utilizando UML e padrões: uma introdução à análise e ao projeto orientados a objetos**. Porto Alegre: Bookman, 2000.

LUNA, José Eduardo Ochoa. **Algoritmos EM para Aprendizagem de Redes Bayesianas a partir de Dados Incompletos**. Disponível em: <http://www.dct.ufms.br/mestrado/dissertacoes/jose_eduardo.pdf>. Acesso em: 25 ago. 2007.

MARQUES, Roberto Ligeiro; DUTRA, Inês. **Redes Bayesianas: o que são, para que servem, algoritmos e exemplos de aplicações**. Disponível em: <<http://www.cos.ufrj.br/~ines/courses/cos740/leila/cos740/Bayesianas.pdf>>. Acesso em: 25 maio 2007.

MURPHY, Kevin. **Pearl's algorithm and multiplexer nodes**. Technical report, U.C. Berkeley, Dept. Comp. Sci., 1999

NASSAR, Silvia Modesto. **Tratamento de Incerteza: Sistemas Especialistas Probabilísticos**. Disponível em: <<http://www.inf.ufsc.br/~silvia/disciplinas/sep/MaterialDidatico.pdf>>. Acesso em: 2 set. 2007.

NEOPOLITAN, Richard E. **Learning in bayesian networks**. New Jersey: Prentice Hall, 2004.

OLIVEIRA JUNIOR, Hime Aguiar e; CALDEIRA, André Machado. **Inteligência computacional aplicada à administração, economia e engenharia em Matlab**. São Paulo: Thomson, 2007.

PEARL, Judea. **Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference**. San Francisco: Morgan Kaufmann, 1988.

REZENDE, Solange Oliveira. **Sistemas inteligentes: fundamentos e aplicações**. Barueri: Manole, 2005.

RUSSEL, Stuart; NORVIG, Peter. **Inteligência artificial**. Rio de Janeiro: Elsevier, 2004.

SCUSSEL, Telma. "Base de Conhecimento para um Sistema Especialista de Apoio ao Diagnóstico Etiológico de Lombalgia". Criciúma, SC: UNESC, 2001. 59 p. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) - Universidade do Extremo Sul Catarinense, 2001.

WANG, Li-xin. **A couse in fuzzy systems and control**. London: Prentice Hall, 1997.

WILLIAMSON, J. **Bayesian nets and causality: philosophical and computational foundations**. New York: Oxford University Press, 2005.

ANEXO A - DIAGNÓSTICO DA LOMBALGIA

“Logo após o resfriado comum, a dor lombar é a aflição mais comum no homem. Entre 70 e 80 % da população do globo sente dor lombar em algum momento de suas vidas” (BERKER,1993).

“A grande maioria dos pacientes com dor lombar tem condições subjacentes que podem ser diagnosticadas e tratadas em ambulatório. A maioria dos pacientes não requer hospitalização ou cirurgia. A tarefa consiste em separar os poucos que exigem um tratamento mais agressivo daqueles que vão se recuperar com manejo conservador” (BERKER,1993).

8.1 DOENÇAS EXISTENTES E DOENÇAS DIAGNOSTICADAS

Existem aproximadamente 50 doenças que cursam com dor lombar , destas, foram contempladas seis, devido a sua alta prevalência .O sistema irá diagnosticar as seis doenças abaixo descritas, bem como os seus sinais e sintomas:

Spondilite Ancilosante: “A espondilite ancilosante (EA) é um distúrbio inflamatório de etiologia desconhecida que acomete principalmente o esqueleto axial, embora também possa atingir articulações periféricas e estruturas extra articulares” (FAUCI,1998).

Artrite (osteoartrite ou osteoartrose): “A osteoartrite, também chamada de artropatia degenerativa, é o tipo mais comum de doença articular. Caracteriza-se pela erosão progressiva da cartilagem articular.

O termo osteoartrite subentende uma doença inflamatória , é considerada uma doença intrínseca da cartilagem na qual alterações bioquímicas e metabólicas

esultam na sua degradação” (COTRAN,2000).

“A osteoartrose é uma das doenças mais antigas e mais comuns do mundo. Trata-se de processo basicamente degenerativo (artrose) e não inflamatório.

A osteoartrose é afecção degenerativa eminentemente crônica das articulações móveis, sobretudo do quadril , joelho e vértebras (que suportam mais peso) e que se caracteriza, basicamente, por degeneração progressiva e perda da cartilagem articular coexistindo esclerose óssea subcondral e proeminências ósseas nas bordas articulares (osteófitos)” (BRASILEIRO,2000).

Artrite Reumatóide: A artrite reumatóide (AR) é uma doença multissistêmica crônica de causa desconhecida.A característica típica da AR é uma sinovite inflamatória, persistente, geralmente comprometendo articulações periféricas em uma distribuição simétrica”(FAUCI,2000).

Espondilolistese: “A espondilolistese é geralmente definida como um deslizamento ou deslocamento anterior ou posterior de uma vértebra, com relação a outra. O defeito unilateral ou bilateral da parte interarticular sem deslocamento da vértebra é conhecido como espondilólise. O termo espondilolistese (foi cunhado por Kilian em 1854)derivado do grego “spondylo”, significando vértebra, e “olisthesis”, significando deslizamento”(CAMPBELL, 1996).

Hérnia Discal: “Discos intervertebrais herniados são a causa mais comum de dor em pescoço ou região lombar associada a uma anormalidade estrutural claramente definida. Entre cada dois corpos vertebrais existe um disco intervertebral fibrocartilaginoso. O disco consiste em núcleo pulposo interno liso (anel fibroso).Um núcleo pulposo gelatinoso age como um absorvente de choque entre corpos vertebrais adjacentes.Com o envelhecimento, o núcleo perde o líquido, volume e elasticidade, e a estrutura do disco torna-se mais suscetível a traumas menores repetidos, e caso

alarguem, uma porção do núcleo pulposo hérnia. Na herniação de disco a maioria dos pacientes mantém o pescoço rígido e resiste a movimentos passivos” (BENNETT,1996).

Lombalgia por causa muscular: Lombalgia por causa muscular é a dor ou desconforto na região lombar, às vezes causada pelo estresse, geralmente envolve a contratura ou espasmo dos músculos ao longo da espinha que servem para sustentar a espinha. Uma dor generalizada ao longo da espinha pode ser devido a estiramento do músculo, trauma nas costas ou estresse mecânico. Lombalgia devido ao estresse pode ser causada pela fadiga ou contração generalizada, devido a permanência numa mesma posição por um tempo prolongado, como trabalhar sentado por horas. O estresse pode deixar a pessoa mais tensa e criar tensão muscular, que é fator de risco da lombalgia.

8.2 SINAIS E SINTOMAS ENVOLVIDOS NO DIAGNÓSTICO

As doenças podem ocorrer na forma clássica, onde os sinais e sintomas permitem ao médico o diagnóstico evidente da patologia, pois esses dificilmente contemplam outras doenças. Porém em grande parte dos casos acontece a interposição de sinais e sintomas caracterizando quadros clínicos não clássicos. Dessa forma, a apresentação clássica das doenças, que permitiria um rápido e fácil diagnóstico infelizmente não aparece corriqueiramente, fazendo com que o conjunto de sinais e sintomas levem o médico geralmente a estabelecer hipóteses diagnósticas de várias doenças, com percentuais probabilísticos diferentes.

Tornando-se essencial para o médico reunir o maior número possível de informações e analisá-las para que possa chegar da forma mais fidedigna a esse diagnóstico. Os sinais e sintomas que compõem o quadro clínico das doenças que

cursam com lombalgia possuem particularidades que à sua análise permitem ao médico chegar com um certo grau de segurança as hipóteses diagnósticas.

Há determinados distúrbios mais comuns em indivíduos mais jovens e outros associados com indivíduos mais idosos, com contribuição relevante relacionada ao sexo. A duração, a forma de início e as características da dor auxiliam na decisão do seguimento e formato da anamnese, sendo direcionada pelo médico no transcorrer da consulta

Além da sintomatologia, fatores como a história ocupacional e social do indivíduo exercem influência na dor lombar. Pacientes jovens, submetidos a esforço muscular exaustivo, sem história de trauma em geral são portadores de lombalgia de causa biomecânica.

Pacientes com hérnia discal classicamente apresentam um quadro de dor irradiada para o membro inferior, associada com déficit sensitivo e motor. Indivíduos idosos, com dor lombar sem irradiação, sem história de trauma em geral são portadores de artrose lombar. Pacientes com quadro de dor poliarticular associada com dor lombar e rigidez matinal possuem maior possibilidade de portarem uma doença reumática. O exame físico tem por objetivo demonstrar anomalias estáticas e dinâmicas que possam descartar ou não determinadas doenças, acrescenta dados essenciais para o diagnóstico sendo condizentes ou não com as queixas referidas pelo paciente durante a anamnese.

Ao final da anamnese e exame físico deve o médico reunir informações necessárias para formular um diagnóstico, conduzindo de forma criteriosa a solicitação de exames adicionais, que confirmem suas hipóteses.

ANEXO A - ARTIGO

O Formalismo Bayesiano pelo Algoritmo de Passagem de Mensagens de Judea Pearl na *Shell Pegasus Uncertainty Modeling*

Edroaldo Lummertz da Rocha¹, Priscyla Waleska Targino de Azevedo Simões¹

¹Curso de Ciência da Computação, Unidade Acadêmica de Ciências, Engenharias e Tecnologia – Universidade do Extremo Sul Catarinense (UNESC) – Criciúma – SC

edroaldo@gmail.com, pri@unesc.net

Abstract. *There are different tools for uncertainty modeling, which use only a knowledge representation mechanism. The objective of this research is the development of a tool, called shell Pegasus Uncertainty Modeling that provides a unified interface different methodologies for uncertainty modeling, as certain factors, fuzzy logic, Dempster-shafer theory and bayesian networks. This paper deals specifically about how bayesian inference is accomplished on it, being the low back pain diagnostic used as problem from which the results are got and analyzed.*

Resumo. *Existem diferentes ferramentas para modelagem da incerteza, as quais utilizam somente um mecanismo de representação de conhecimento. O objetivo desta pesquisa é o desenvolvimento de uma ferramenta, intitulada shell Pegasus, que propicie em uma interface unificada diferentes metodologias para modelagem da incerteza, como fatores de certeza, lógica fuzzy, teoria de Dempster-Shafer e redes bayesianas. Este artigo trata especificamente sobre como a inferência bayesiana é realizada na mesma, sendo o diagnóstico de lombalgia utilizado como problema a partir do qual os resultados são obtidos e analisados.*

1. Introdução

Na área de inteligência artificial existem diferentes ferramentas para modelagem da incerteza, dentre as quais estão a *shell Netica*, para redes bayesianas, a *shell Expert Sinta*, para fatores de certeza, o *toolbox fuzzy* do Matlab e a Dempster-shafer engine para a teoria de Dempster-Shafer. Estas teorias abordam variações do conceito de incerteza e estas ferramentas automatizam o desenvolvimento de sistemas inteligentes, fornecendo uma interface amigável para utilização destas abordagens para modelá-la.

No entanto, estas ferramentas não abordam a disponibilização destas teorias em uma única ferramenta e neste sentido, o Grupo de Pesquisa em Inteligência Computacional Aplicada da Universidade do Extremo Sul Catarinense propôs o desenvolvimento de uma ferramenta, denominada *shell Pegasus Uncertainty Modeling*, cujo objetivo é fornecer um ambiente integrado de desenvolvimento e avaliação dos modelos criados, baseada inicialmente, nas quatro teorias acima citadas.

Os objetivos a longo prazo dividem-se em duas categorias: globais e locais. Os objetivos globais caracterizam a integração da teoria de redes bayesianas com as teorias de lógica *fuzzy* e redes neurais artificiais, proporcionando um mecanismo para criação de abordagens híbridas para modelagem de sistemas inteligentes. Os objetivos locais

definem o aprofundamento tanto matemático quanto algorítmico dos aspectos relacionados às respectivas teorias. Por exemplo, tratando-se da teoria de redes bayesianas, abordar o desenvolvimento de algoritmos de inferência exata, aproximada e simbólica bem como os de aprendizado, subdivididos nos métodos de independência condicional e busca e pontuação. Atualmente, esta ferramenta disponibiliza mecanismos de inferência para fatores de certeza, Dempster-Shafer e redes bayesianas, os quais encontram-se em funcionamento.

Neste sentido, este artigo apresenta uma descrição da *shell* Pegasus bem como a implementação do módulo correspondente a teoria de redes bayesianas, especificamente o algoritmo de passagem de mensagens (PEARL, 1988) em grafos com estrutura de poliárvore.

1.1. Redes Bayesianas

Redes bayesianas são grafos acíclicos dirigidos constituídos por dois componentes fundamentais: a) um componente estrutural, S , que define um relacionamento qualitativo causal entre os nós de um grafo, e b) parâmetros numéricos, θ , que quantificam a relação probabilística causal existente entre os nós de S (LUNA, 2004) (WILLIANSO, 2005).

Considerando que em um relacionamento causal a causa precede o efeito, há a necessidade de um mecanismo unidirecional para modelagem da causalidade, sendo o conceito de probabilidade condicional indicativo desses relacionamentos sem conexão direta (NASSAR, 2003).

Neste sentido, as redes bayesianas fornecem um modelo matemático que expressa causalidade explorando as relações de independência entre as variáveis no domínio sendo modelado e possibilitando a fatorização da distribuição de probabilidade conjunta. Esta distribuição permite responder a diversas questões sobre um domínio de dados. No entanto, sua dificuldade de representação aumenta em função do número de variáveis.

Na aplicação do teorema de Bayes, a exploração de independências condicionais entre variáveis aleatórias possibilita simplificar os cálculos e reduzir o número de probabilidades condicionais que precisam ser especificadas. Sob este aspecto, as redes bayesianas, utilizando o teorema de Bayes como mecanismo de inferência, representam as dependências entre as variáveis aleatórias existentes bem como fornecem uma representação compacta da distribuição de probabilidade conjunta, caracterizando uma estrutura eficiente para a realização do procedimento de inferência (PEARL, 1988)

1.2. Inferência

A inferência bayesiana possibilita estimar a probabilidade das variáveis aleatórias pertencentes a um domínio de aplicação modelado a partir da ocorrência de um conjunto de evidências E (CASTILHO; GUTIÉRREZ; HADI, 1998). Sua importância está relacionada não somente a obtenção de distribuições de probabilidade condicionais como também para aprendizagem automática de redes bayesianas.

Em contextos médicos, por exemplo, o diagnóstico para determinada doença é obtido analisando-se os sintomas (evidências) apresentados pelo paciente. Este mecanismo é denominado propagação de evidências (inferência) e consiste em atualizar as probabilidades das variáveis quando novas informações são fornecidas.

Existem diferentes algoritmos de inferência, os quais podem ser classificados

em três categorias fundamentais (NEOPOLITAN, 2004): algoritmos exatos, aproximados e simbólicos. Os algoritmos exatos calculam as probabilidades das variáveis sem outro erro senão aquele produzido pelas limitações de representação numérica do processador (CASTILHO; GUTIÉRREZ; HADI, 1998).

No entanto, em situações onde os métodos exatos não são aplicáveis, utilizam-se algoritmos aproximados, os quais aplicam diferentes técnicas de simulação para obter probabilidades aproximadas. Os algoritmos simbólicos podem ser utilizados tanto com parâmetros numéricos quanto simbólicos, obtendo probabilidades na forma simbólica em função dos parâmetros (LUNA, 2004).

A efetivação de inferência bayesiana na *shell* Pegasus é realizada com o algoritmo de passagem de mensagens apresentado em (PEARL, 1988), em sua versão para poliárvores. Em uma poliárvore, dois nós estão conectados por apenas um caminho, fazendo com que cada nó subdivida esta estrutura em duas poliárvores desconexas, uma contendo os pais e outra os filhos do respectivo nó, além dos subgrafos associados a cada pai e filho do nó em questão (CASTILHO; GUTIÉRREZ; HADI, 1998), conforme ilustra a Figura 3.

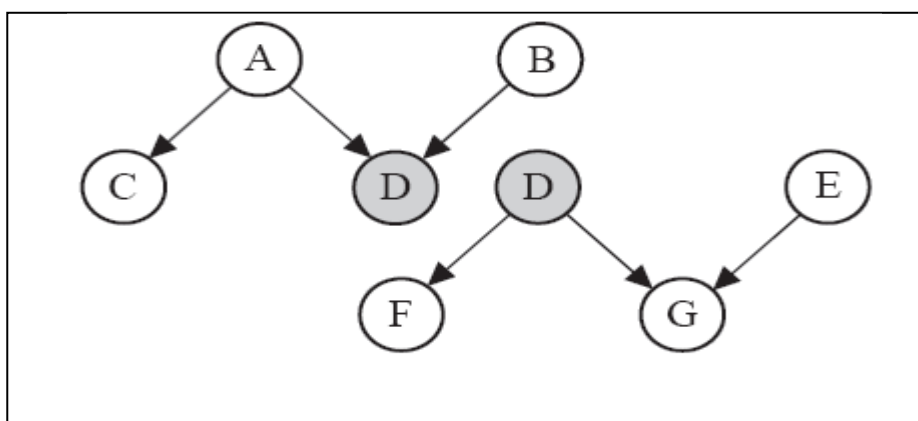


Figura 3. O nó D divide a poliárvore em duas poliárvores desconexas
 Fonte: CASTILHO, E.; GUTIÉRREZ, J.; HADI, A. (1998).

Por exemplo, na Figura 3, o nó D divide a poliárvore em duas poliárvores não conexas. A primeira, constituída pelos nós {A, B, C}, inclui seus pais e os nós acessíveis desde D por meio destes. A segunda, formada pelos nós {E, F, G}, inclui seus filhos e os nós acessíveis desde D por meio de seus filhos. Conforme ilustrado, o nó D separa estes dois conjuntos e, portanto, a relação de independência condicional $I(\{A, B, C\}\{E, F, G\}|D)$ é graficamente constatada.

Neste tipo de grafo, a inferência bayesiana pode ser eficientemente realizada utilizando passagens de mensagens²¹ entre os distintos subgrafos da poliárvore, por meio de computações locais²² (NEOPOLITAN, 2004). Portanto, o algoritmo de passagens de mensagens (PEARL, 1988) pode ser utilizado para propagar as probabilidades pela rede quando um conjunto de evidências E é fornecido. A seguir é descrito um resumo deste algoritmo.

²¹ Mecanismo pelo qual as probabilidades são distribuídas pela rede bayesiana (NEOPOLITAN, 2004).

²² Cada nó que compõe a rede faz cálculos localmente, baseados em sua informação e na informação proveniente de outros nós, que uma vez calculada é transmitida para outros nós, por meio de uma mensagem (KORB; NICHOLSON, 2004).

O resumo abaixo descrito é baseado em (KORB; NICHOLSON, 2004). Considerando uma variável X , com m filhos e n pais. A distribuição de probabilidades da variável X pode ser calculada se três tipos de parâmetros estão disponíveis:

- a) A força atual do suporte causal, π , formado pela contribuição de cada arco de chegada $U_i \rightarrow X$:

$$\pi_X(x_i) = P(u_i | e_{U_i X}^+)$$

- b) A força atual do suporte diagnóstico, λ , formado pela contribuição de arco de saída $X \rightarrow Y_j$:

$$\lambda_{Y_j}(x) = P(e_{XY_j}^- | x)$$

- c) A matriz de probabilidade condicional fixa $P(x | u_1, \dots, u_n)$ que relaciona a variável X a seus pais imediatos.

Utilizando estes parâmetros, a atualização de crença local pode ser executada em três passos:

Passo 1- Atualização de crença: quando um nó X é ativado (recebe uma evidência), ele simultaneamente inspeciona as mensagens $\pi_X(x_i)$ enviadas pelos seus pais e as mensagens $\lambda_{Y_j}(x)$ enviadas pelos seus filhos. Utilizando esta entrada, o nó atualiza sua medida de crença para:

$$BEL(x) = \alpha \lambda(x) \pi(x)$$

Onde:

$$\lambda(x) = \prod_{j=1}^m \lambda_{Y_j}(x)$$

$$\pi(x) = \sum_{\mathbf{u}} P(x | \mathbf{u}) \prod_{i=1}^n \pi_X(u_i)$$

Passo 2- Propagação ascendente: Utilizando as mensagens recebidas, o nó X calcula novas mensagens λ para serem enviadas aos seus pais. Por exemplo, a nova mensagem $\lambda_X(u_i)$ que X envia a seus pais U_i é calculada por:

$$\lambda_X(u_i) = \beta \sum_x \lambda(x) \sum_{u_k: k \neq i} P(x | \mathbf{u}) \prod_{k \neq i} \pi_X(u_k)$$

Passo 3- Propagação descendente: Cada nó calcula novas mensagens π para serem enviadas aos seus filhos. Por exemplo, a nova mensagem $\pi_{Y_j}(x)$ que X envia aos seus filhos Y_j é calculada por:

$$\pi_{Y_j}(x) = \alpha \left[\prod_{k \neq j} \lambda_{Y_k}(x) \right] \sum_{\mathbf{u}} P(x | \mathbf{u}) \prod_{i=1}^n \pi_X(u_i)$$

$$\pi_{Y_j}(x) = \alpha \frac{BEL(x)}{\lambda_{Y_j}(x)}$$

2. Shell Pegasus

A *shell* Pegasus é uma ferramenta para modelagem da incerteza constituída por quatro módulos principais: teoria dos fatores de certeza, lógica *fuzzy*, teoria de Dempster-shafer e redes bayesianas. O objetivo de seu desenvolvimento é prover um ambiente unificado para análise de abordagens de construção de sistemas inteligentes e metodologias híbridas para modelagem de problemas baseados em conhecimento incerto.

Ela está sendo desenvolvida com a tecnologia Java e o ambiente de Netbeans 6.1, sendo sua modelagem realizada com um *plugin* UML integrado ao próprio ambiente de desenvolvimento. A Figura 2 ilustra a interface da *shell* Pegasus com o módulo de redes bayesianas inicializado.

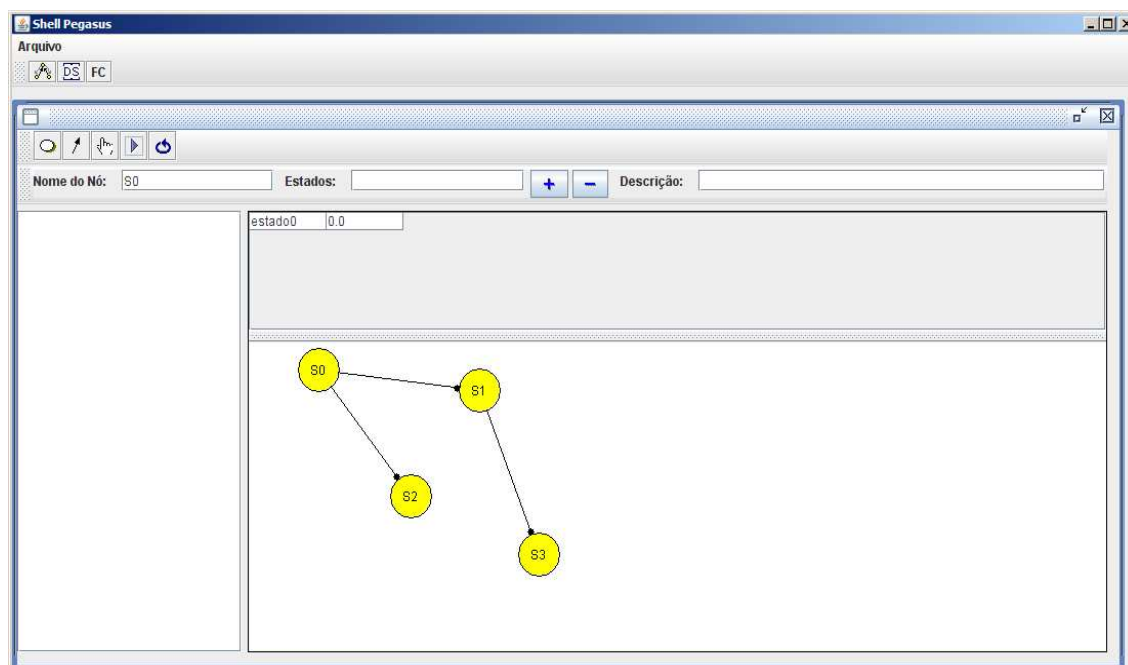


Figura 2. Interface da *shell* Pegasus com o módulo bayesiano ativado.

A metodologia abordada neste artigo consiste na apresentação do módulo de redes bayesianas, da base de conhecimento e dos testes realizados no módulo de redes bayesianas.

3. Metodologia

Primeiramente, foi criada uma base de conhecimento na *shell* Pegasus, com o módulo redes bayesianas, utilizando a base desenvolvida em Scussel (2001) para obter os relacionamentos e distribuições de probabilidade associadas ao diagnóstico de lombalgia, a partir da qual foram realizados testes e obtidos os resultados.

3.1. Módulo de Redes Bayesianas

As características relacionadas a redes bayesianas são definidas neste módulo, o qual é responsável por agrupar desenvolvimentos futuros como novos algoritmos de inferência e aprendizado. Em seu estágio inicial, este módulo é constituído por um mecanismo de inferência baseado em troca de mensagens (NEOPOLITAN, 2004) entre os nós que constituem uma rede bayesiana, apresentado em Pearl (1988) e descrito na seção 2.1.

Utilizando este algoritmo, foram realizados testes aplicando-o ao problema do diagnóstico etiológico de lombalgia, com uma base de conhecimento desenvolvida em Scussel (2001), a qual foi escolhida objetivando a comparação dos resultados obtidos pela *shell* Netica, ferramenta na qual a base de (Scussel, 2001) foi desenvolvida, e aqueles obtidos pelo módulo de redes bayesianas da *shell* Pegasus.

3.2. Base de Conhecimento

A base de conhecimento utilizada nesta pesquisa foi desenvolvida em 2001 como um trabalho de conclusão de curso na UNESC e uma descrição detalhada da mesma encontra-se em (SCUSSEL, 2001).

Esta base de conhecimento, ilustrada na Figura 1, foi criada utilizando a *shell* Netica sob orientação de um médico especialista em traumatologia, do curso de medicina da UNESC. Ela é composta por 23 nós, sendo as hipóteses diagnósticas especificadas no nó Lombalgia e os demais nós são denominados nós evidenciais.

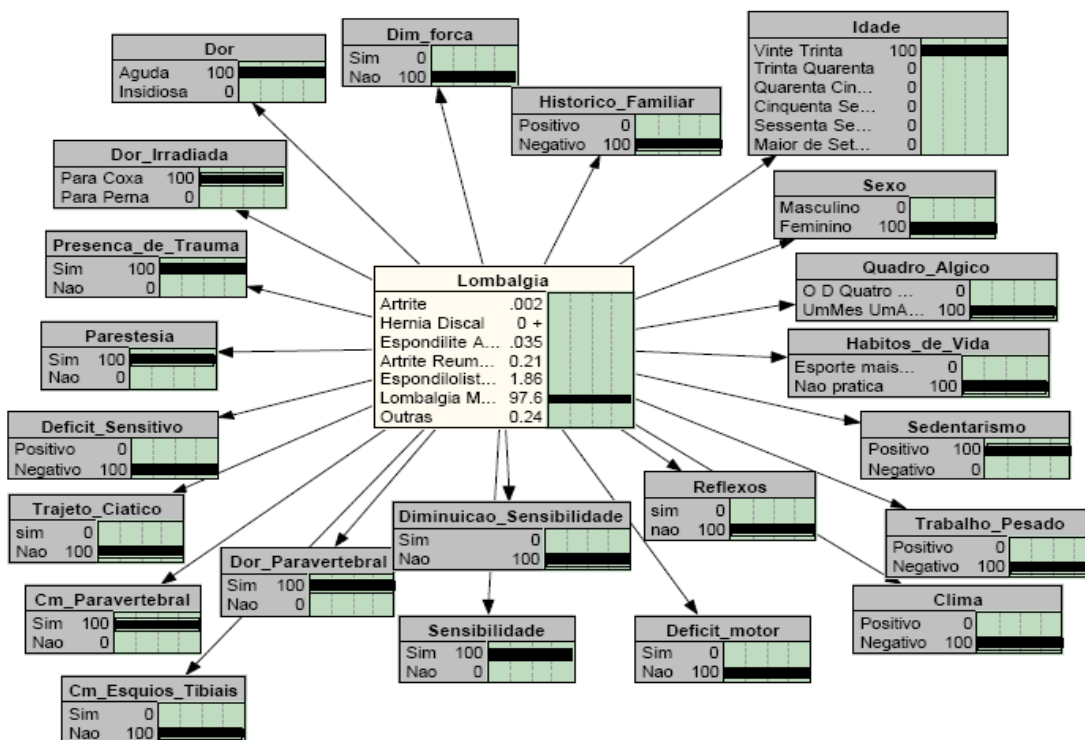


Figura 1. Base de conhecimento para diagnóstico de lombalgia.

Fonte: SCUSSEL, T. (2001)

A partir das tabelas de probabilidade condicionais e *a priori* e dos relacionamentos existentes entre os nós que constituem esta base de conhecimento, uma base quantitativa e qualitativamente equivalente foi criada na *shell* Pegasus.

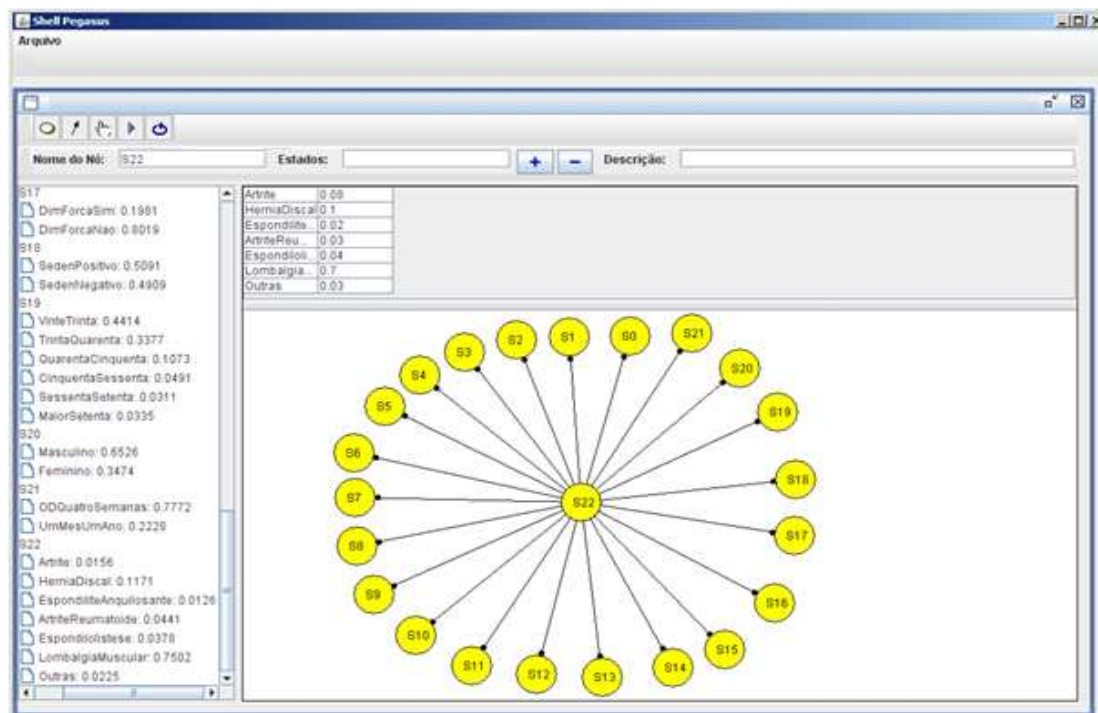


Figura 4. Interface da *shell* Pegasus com o módulo bayesiano ativado.

O nó S22 corresponde ao nó de hipóteses diagnósticas (Lombalgia) e os demais são os nós evidenciais. A partir da construção deste modelo, foram realizadas inferências com diferentes tipos de entradas, sendo os mesmos procedimentos empregados na base de conhecimento desenvolvida na *shell* Netica.

4. Resultados

A partir da metodologia apresentada na seção 5, os resultados numéricos provenientes da *shell* Pegasus mostraram-se condizentes com os obtidos pela *shell* Netica, demonstrando um comportamento estável do algoritmo de inferência desenvolvido. É importante ressaltar que este algoritmo funciona corretamente somente com redes bayesianas que apresentam estrutura de poliárvore, uma vez que o mesmo não é projetado para redes bayesianas multiconectadas.

Os valores de probabilidade do nó Lombalgia (S22), quando nenhuma evidência é fornecida são ilustrados na Figura 5.

	0.08	Lombalgia	
Artrite	0.08	Artrite	8.00
HerniaDiscal	0.1	Hernia Discal	10.0
Espondilite...	0.02	Espondilite Anquilosante	2.00
ArtriteReu...	0.03	Artrite Reumatoide	3.00
Espondiloli...	0.04	Espondilolite	4.00
Lombalgia ...	0.7	Lombalgia Muscular	70.0
Outras	0.3	Outras	3.00

Figura 5. (a) Resultados obtidos na *shell* Pegasus (b) Resultados obtidos na *shell* Netica

Uma evidência é fornecida clicando duas vezes no estado da respectiva variável. Automaticamente, o estado selecionado assume o valor 1 enquanto os demais são configurados para zero e a distribuição de probabilidades do modelo é atualizada refletindo a ocorrência da evidência. A Figura 6 apresenta as probabilidades do nó Lombalgia (S22) quando a evidência Histórico Familiar (S1) assume o estado Positivo.

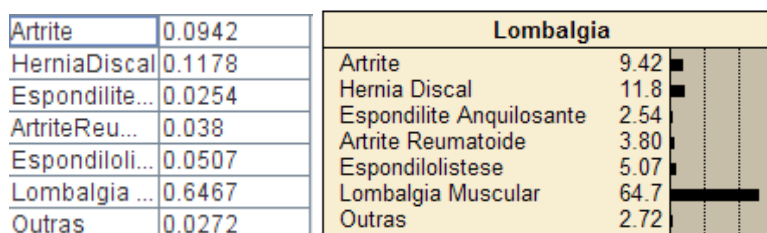


Figura 6. (a) Resultados obtidos na *shell* Pegasus (b) Resultados obtidos na *shell* Netica

Conforme ilustra a Figura 6, a ocorrência da evidência Histórico Familiar positivo aumenta a probabilidade da ocorrência de Lombalgia Muscular. Esta informação, em um processo decisório, poderia auxiliar o médico na realização de procedimentos específicos relacionados a este estado ao invés dos demais estados.

Em seguida, fornecendo a evidência Dor no estado Aguda, as probabilidades da rede bayesiana são atualizadas, tanto para a evidência Histórico Familiar assumindo o estado Positivo, quanto para Dor em seu estado Aguda, as quais são ilustradas na Figura 7.



Figura 7. (a) Resultados obtidos na *shell* Pegasus (b) Resultados obtidos na *shell* Netica

Conforme ilustra a Figura 7, a ocorrência das evidências Histórico Familiar positivo e Dor Aguda aumentam ainda mais a probabilidade da ocorrência de Lombalgia Muscular.

5. CONSIDERAÇÕES FINAIS

O modelo bayesiano desenvolvido utilizando a *shell* Pegasus apresenta resultados condizentes com aqueles apresentados pela *shell* Netica, portanto, se constata o adequado funcionamento do algoritmo de inferência quando a rede bayesiana apresenta uma estrutura de poliárvore.

Considerando os relacionamentos existentes entre as variáveis e a atualização da distribuição de probabilidades como um reflexo do fornecimento de evidências constata-se que as redes bayesianas são uma mecanismo intuitivo para auxílio a tomada de decisão, especialmente em contextos médicos, onde o raciocínio realizado em processos diagnósticos pode ser adequadamente representado por estes modelos.

Referências

PEARL, Judea. **Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference**. San Francisco: Morgan Kaufmann, 1988.

SCUSSEL, Telma. "Base de Conhecimento para um Sistema Especialista de Apoio ao Diagnóstico Etiológico de Lombalgia". Criciúma, SC: UNESC, 2001. 59 p. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) - Universidade do Extremo Sul Catarinense, 2001.

NEOPOLITAN, Richard E. **Learning in bayesian networks**. New Jersey: Prentice Hall, 2004.

LUNA, José Eduardo Ochoa. **Algoritmos EM para Aprendizagem de Redes Bayesianas a partir de Dados Incompletos**. Disponível em:

<http://www.dct.ufms.br/mestrado/dissertacoes/jose_eduardo.pdf>. Acesso em: 25 ago. 2007.

WILLIAMSON, J. **Bayesian nets and causality: philosophical and computational foundations**. New York: Oxford University Press, 2005.

NASSAR, Silvia Modesto. **Tratamento de Incerteza: Sistemas Especialistas Probabilísticos**. Disponível em:

<<http://www.inf.ufsc.br/~silvia/disciplinas/sep/MaterialDidatico.pdf>>. Acesso em: 2 set. 2007.

CASTILHO, Enrique; GUTIÉRREZ, José Manuel; HADI, Ali S. **Sistemas Expertos y Modelos de Redes Probabilísticas**. Madrid: Academia Española de Ingeniería, 1998.

KORB, Kevin; NICHOLSON, Ann. **Bayesian artificial intelligence**. Florida: Chapman & Hall, 2004.

APÊNDICE A – MODELAGEM MATEMÁTICA DO ALGORITMO DE INFERÊNCIA

Este apêndice descreve de uma forma alternativa o exemplo numérico realizado no capítulo 7. Como no exemplo, os nós representam as variáveis Febre, Dor de Cabeça e Gripe e considera-se que cada nó possui apenas dois estados. A Figura 34 ilustra o modelo criado para o exemplo e a Tabela 10 apresenta o nome dos nós, seus respectivos estados e significados.

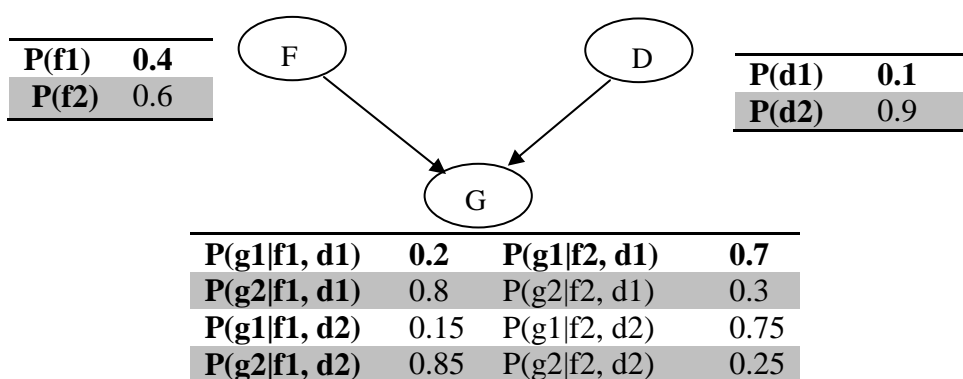


Figura 34. Rede usada no exemplo

Tabela 10. Variáveis, estados e seus significados

Var	Estado	Significado
F	f1	Possui febre
	f2	Não possui febre
D	d1	Está com dor no corpo
	d2	Não está com dor no corpo
G	g1	Tem gripe
	g2	Não tem gripe

Para inicializar a rede bayesiana realiza-se uma chamada ao método

Initial_tree(), resultando nos seguintes cálculos:

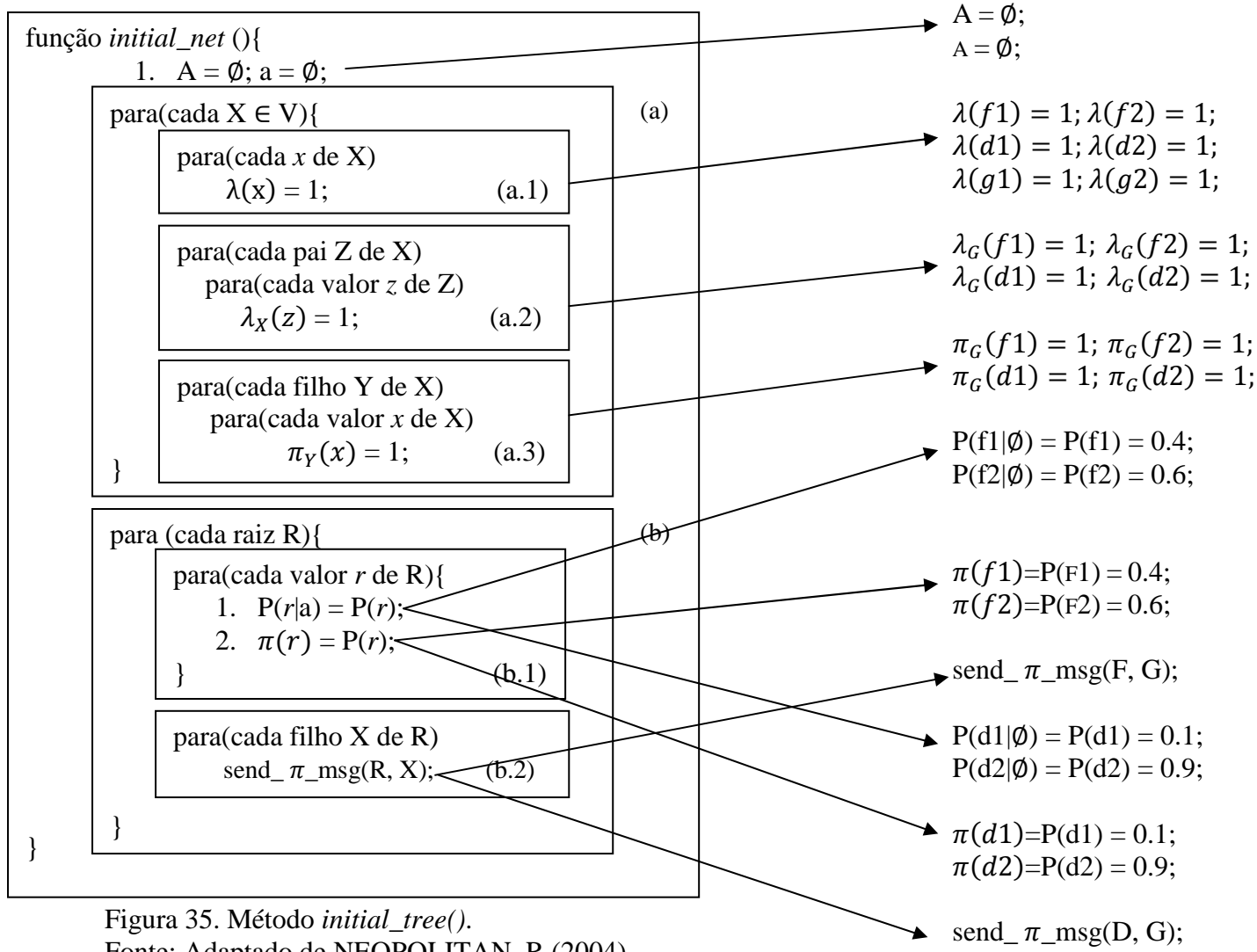


Figura 35. Método *initial_tree*().
 Fonte: Adaptado de NEOPOLITAN, R.(2004)

```

função send_pi_msg(Nó Z, Nó X){
  para(cada valor de z)
    
$$\pi_x(z) = \pi(z) \prod_{Y \in CH_z - \{X\}} \lambda_Y(z);$$

  se (X ∉ A){
    para(cada valor de x){
      
$$\pi(x) = \sum_{z_1, z_2, \dots, z_j} \left( P(x|z_1, z_2, \dots, z_j) \prod_{i=1}^j \pi_X(z_i) \right); \quad (b.1)$$

      
$$P(x|a) = \alpha \lambda(x) \pi(x); \quad (b.2)$$

    }
    normalizar P(x|a);
    para(cada filho Y de X)
      send_pi_msg(X, Y);
  }
  1. se não( $\lambda(x) = 1$  para todos os valores de x)
  2. para(cada pai W de X tal que  $W \neq Z$  e  $W \notin A$ )
  3. send_lambda_msg(X, W);
}

```

A chamada send_pi_msg(F, G) resulta nos seguintes valores:

(a) $\pi_G(f1) = \pi(f1) = 0.4;$
 $\pi_G(f2) = \pi(f2) = 0.6;$

(b) $\pi(g1) =$
 $P(g1|f1, d1)\pi_G(f1)\pi_G(d1) + P(g1|f1, d2)\pi_G(f1)\pi_G(d2)$
 $+ P(g1|f2, d1)\pi_G(f2)\pi_G(d1) + P(g1|f2, d2)\pi_G(f2)\pi_G(d2)$
 $\pi(g1) = (0.2)(0.4)(1) + (0.15)(0.4)(1) + (0.7)(0.6)(1) + (0.75)(0.6)(1)$
 $\pi(g1) = 1.01$

$\pi(g2) =$
 $(g2|f1, d1)\pi_G(f1)\pi_G(d1) + P(g2|f1, d2)\pi_G(f1)\pi_G(d2)$
 $+ P(g2|f2, d1)\pi_G(f2)\pi_G(d1) + P(g2|f2, d2)\pi_G(f2)\pi_G(d2)$
 $\pi(g2) = (0.8)(0.4)(1) + (0.85)(0.4)(1) + (0.3)(0.6)(1) + (0.25)(0.6)(1)$
 $\pi(g2) = 0.99$

$P(g1|\emptyset) = \alpha \lambda(g1) \pi(g1) = \alpha(1)(1.01) = 1.01\alpha$
 $P(g2|\emptyset) = \alpha \lambda(g2) \pi(g2) = \alpha(1)(0.99) = 0.99\alpha$

(c) $P(g1|\emptyset) = \frac{1.01\alpha}{1.07\alpha + 0.99\alpha} = 0.505$
 $P(g2|\emptyset) = \frac{0.99\alpha}{1.01\alpha + 0.99\alpha} = 0.495$

Figura 36. Método send_pi_msg().
 Fonte: Adaptado de NEOPOLITAN, R.(2004)

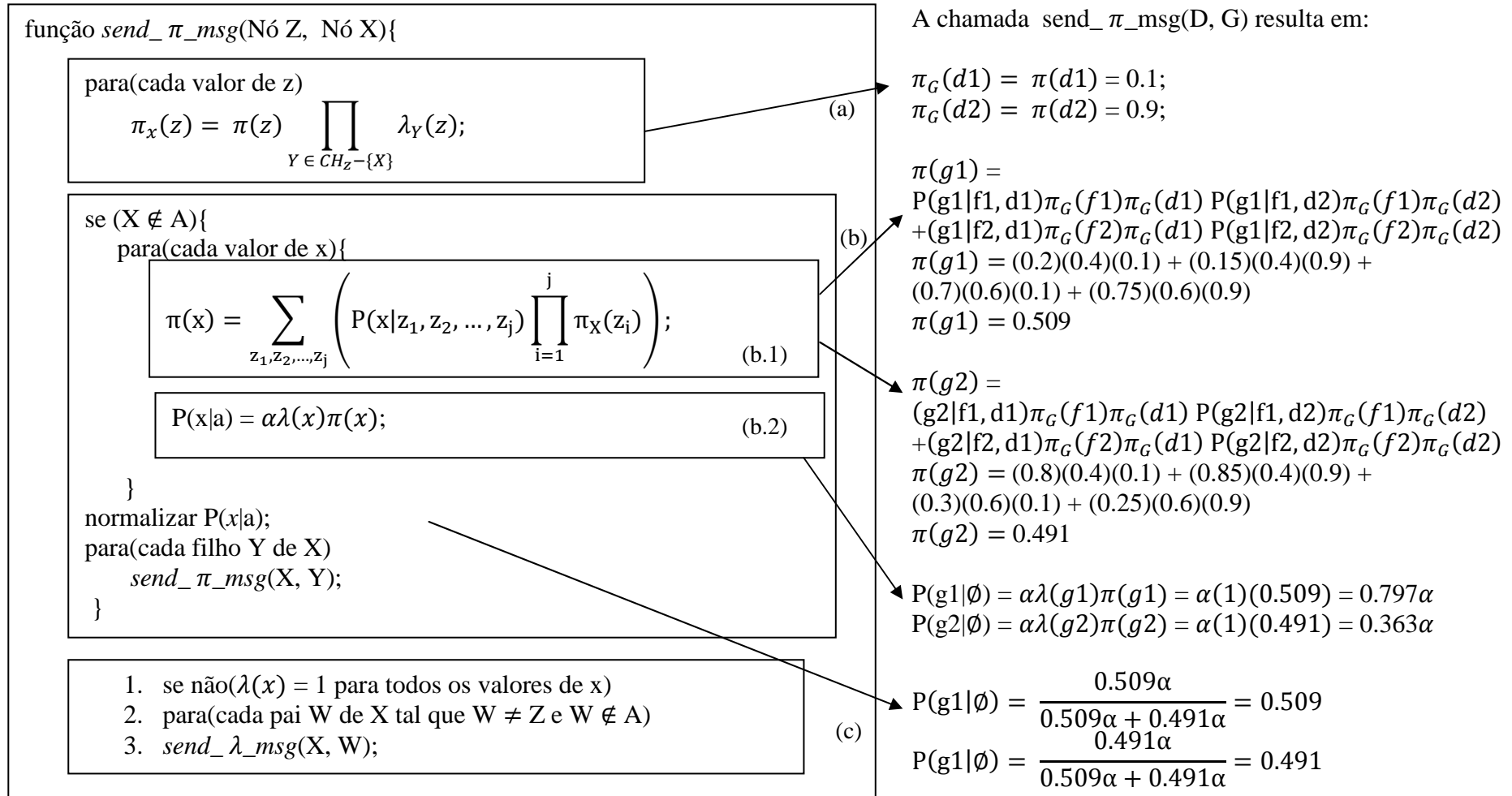


Figura 37. Método $send_pi_msg()$.
 Fonte: Adaptado de NEOPOLITAN, R.(2004)

O fornecimento da evidência g_1 (tem gripe) é refletido propagando-se as probabilidades pela rede bayesiana e é executado utilizando uma chamada ao método *update_tree()*, o que resultada nos seguintes cálculos:

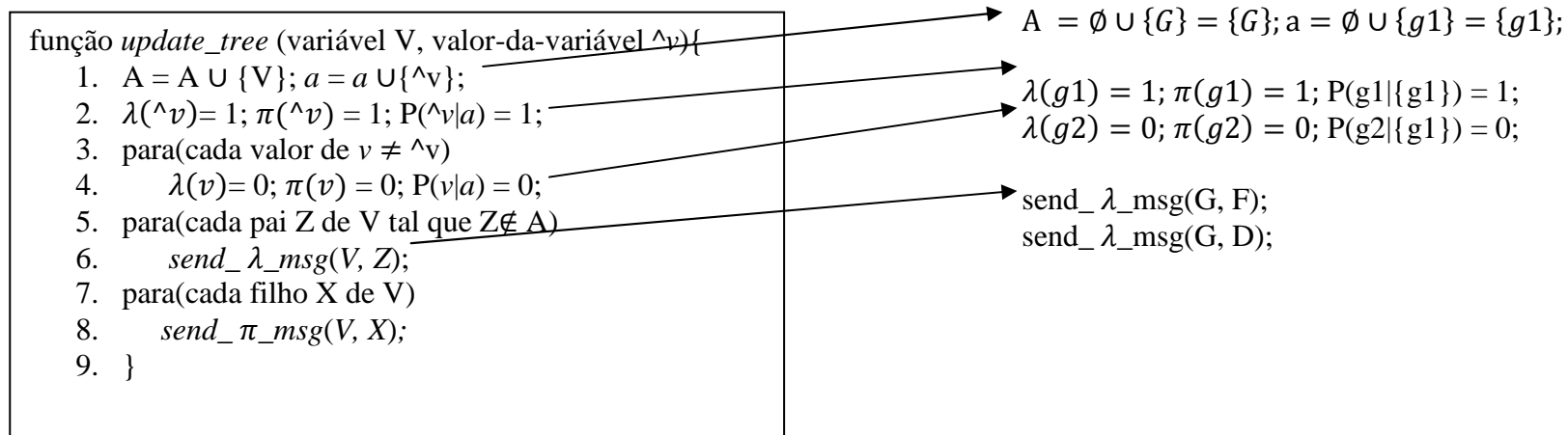
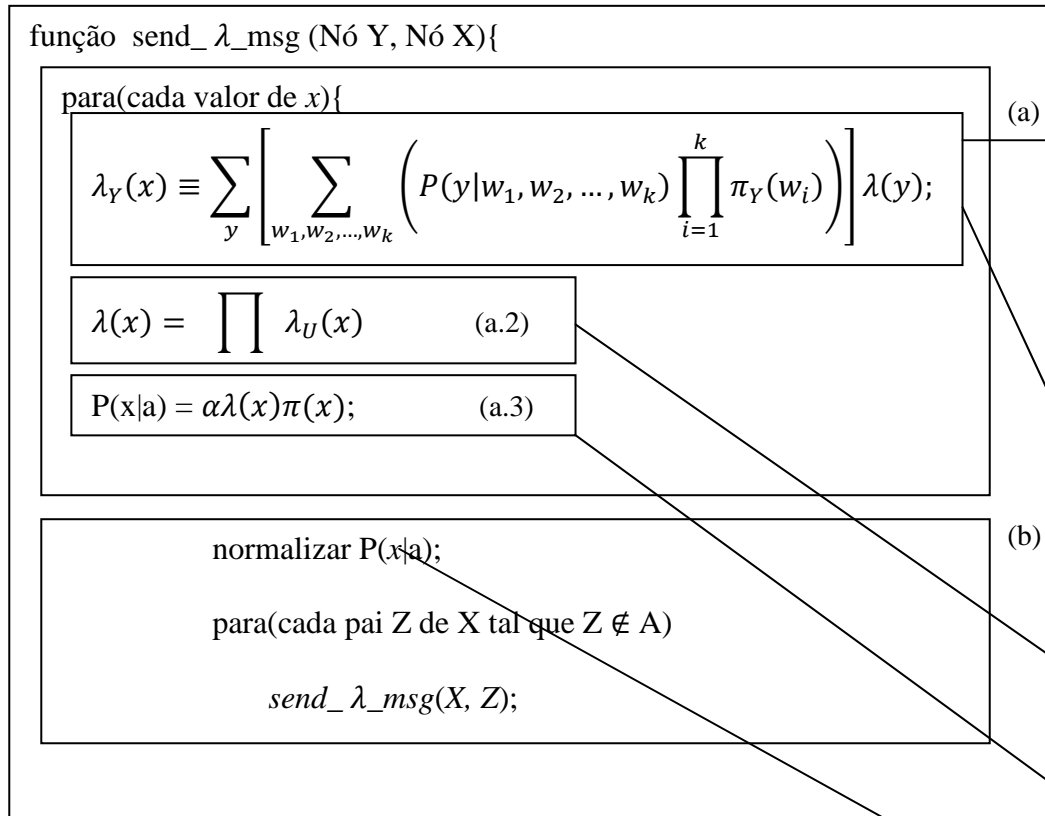


Figura 38. Método *update_tree()*
Fonte: Adaptado de NEOPOLITAN, R.(2004)



A chamada *send_λ_msg*(G, F) resulta em:

$$\lambda_G(f1) = [P(g1|f1, d1)\pi_G(d1) + P(g1|f1, d2)\pi_G(d2)]\lambda(g1) + [P(g2|f1, d1)\pi_G(d1) + P(g2|f1, d2)\pi_G(d2)]\lambda(g2)$$

$$\lambda_G(f1) = [(0.2)(0.1) + (0.15)(0.9)]1 + [(0.8)(0.1) + (0.85)(0.9)]0$$

$$\lambda_G(f1) = 0.155$$

$$\lambda_G(f2) = [P(g1|f2, d1)\pi_G(d1) + P(g1|f2, d2)\pi_G(d2)]\lambda(g1) + [P(g2|f2, d1)\pi_G(d1) + P(g2|f2, d2)\pi_G(d2)]\lambda(g2)$$

$$\lambda_G(f2) = [(0.7)(0.1) + (0.75)(0.9)]1 + [(0.3)(0.1) + (0.25)(0.9)]0$$

$$\lambda_G(f2) = 0.745$$

$$\lambda(f1) = \lambda_G(f1) = 0.155;$$

$$\lambda(f2) = \lambda_G(f2) = 0.745;$$

$$P(f1|\{g1\}) = \alpha \lambda(f1) \pi(f1) = \alpha(0.155)(0.4) = 0.062\alpha$$

$$P(f2|\{g1\}) = \alpha \lambda(f2) \pi(f2) = \alpha(0.745)(0.6) = 0.447\alpha$$

$$P(f1|\{g1\}) = \frac{0.062\alpha}{0.062\alpha + 0.447\alpha} = 0.1218$$

$$P(f2|\{g1\}) = \frac{0.447\alpha}{0.062\alpha + 0.447\alpha} = 0.8781$$

Figura 39. Método *send_λ_msg*(*.*)
 Fonte: Adaptado de NEOPOLITAN, R.(2004)

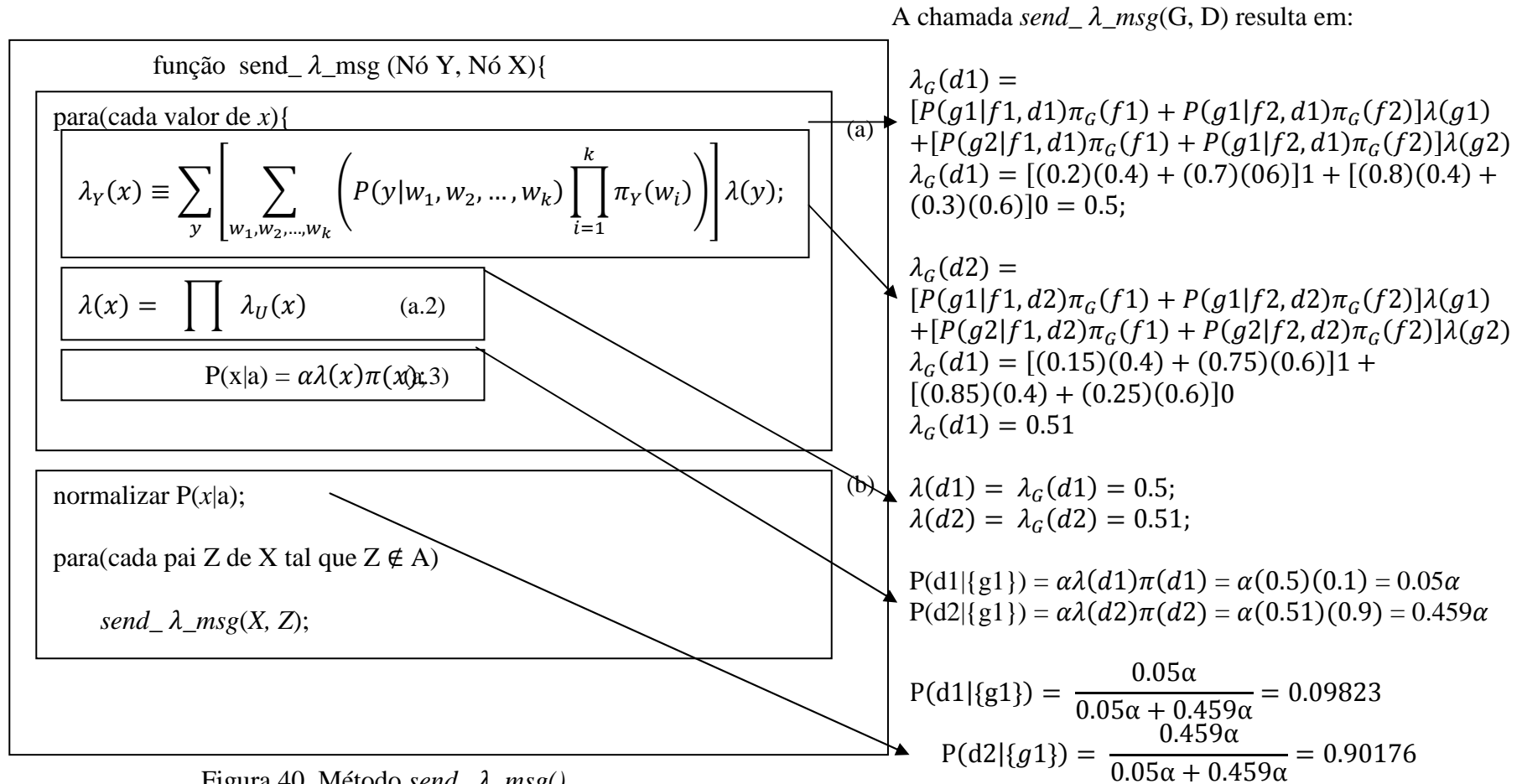


Figura 40. Método *send_λ_msg*().
 Fonte: Adaptado de NEOPOLITAN, R.(2004)

