

ISAQUE ALVES DE LYRA

INTFORBAYES:
UMA FERRAMENTA PARA SISTEMAS ESPECIALISTAS BAYESIANOS

Dissertação apresentada ao Curso de
MESTRADADO EM INFORMATICA da
Universidade Federal da Paraíba,
em cumprimento às exigências para
obtenção do Grau de Mestre.

HELIO DE MENEZES SILVA
orientador

Dissertação (012)
L. 012



L992i Lyra, Isaque Alves de
Intforbayes : uma ferramenta para sistemas especialistas bayesianos / Isaque Alves de Lyra. - Campina Grande, 1991. 114 f. : il.

Dissertacao (Mestrado em Informatica) - Universidade Federal da Paraiba, Centro de Ciencias e Tecnologia.

1. Sistemas Especialistas - 2. Sistemas Especialistas Bayesianos 3. Dissertacao I. Silva, Helio de Menezes, M.Sc. II. Universidade Federal da Paraiba - Campina Grande (PB) III. Título

CDU 004.891(043)

IntForBayes: UMA FERRAMENTA PARA SISTEMAS ESPECIALISTAS
BAYESIANOS

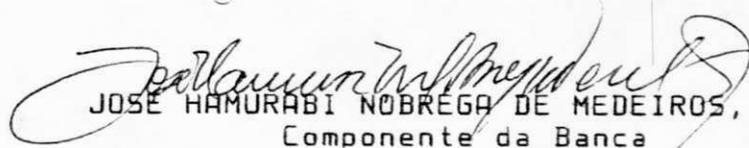
ISAQUE ALVES DE LYRA

DISSERTAÇÃO APROVADA EM 17.07.1991


HELIO DE MENEZES SILVA, M.Sc
Orientador


MISAEEL ELIAS DE MORAIS, Dr.
Componente da Banca


GIUSEPPE MONGIOVI, M.Sc
Componente da Banca


JOSE HAMURABI NOBREGA DE MEDEIROS, M.Sc
Componente da Banca

Campina Grande, 17 de julho de 1991

AGRADECIMENTOS

Ao orientador Hélio de Menezes Silva pela valiosa amizade, pelo seu grande incentivo e pela sua extraordinária dedicação expressada em todas as fases de desenvolvimento deste trabalho.

Aos meus pais pelos incentivos proporcionados ao longo de toda a minha vida de estudante.

A Andréia Fernandes pela digitação desta dissertação.

Aos professores João Damasco e José Hamurabi, pelas valiosas idéias e contribuições.

Ao CNPq e BNB pelos apoios financeiros.

A Dra. Paula Franssinetti Vasconcelos de Medeiros por revisar a terminologia médica.

Aos professores Hélio de Menezes, Misael Elias, Giuseppe Mongiovi e José Hamurabi pela composição da banca examinadora desta dissertação.

Ao professor Brasileiro e aos alunos Raquel, Rostand, Walfredo, Tojal, Tércio e Plínio pela implementação de grande parte do editor de IntForBayes.

A Aurélio, Odete e Andréia.

INTFORBAYES
UMA FERRAMENTA PARA SISTEMAS ESPECIALISTAS BAYESIANOS

RESUMO

Este trabalho descreve IntForBayes, um "shell" para desenvolvimento e consulta de sistemas especialistas, cujas principais vantagens são: inteligência na determinação da ordem e número das perguntas a serem feitas (esta inteligência corresponde a uma "best-first search", com valores heurísticos calculados localmente, refletindo a importância que os fatos têm para provar/desprovar a hipótese escolhida); raciocínio com encadeamento podendo ser escolhido entre 100% progressivo ("forward") até 100% regressivo ("backward"); inferência probabilística baseada no teorema de Bayes, lidando com incerteza e incompletude; poda do espaço de pesquisa; parada antecipada do ciclo de investigação de uma hipótese ao ser alcançado um ponto de satisfação na definição da sua plausibilidade; e a possibilidade de "voluntariamento" de informações não solicitadas pelo sistema. O uso combinado e inteligente de todas estas características pode ser bastante vantajoso.

ABSTRACT

This work describes IntForBayes, a shell for developing and consulting expert systems. Its main advantages are: intelligence in the determination of the order and number of questions to be asked (this intelligence corresponds to a best-first search whose heuristic values are locally calculated, reflecting the importance of the facts for proving/refuting the chosen hypothesis); chain of reasoning chooseable between purely forward and purely backward; probabilistic inference based on Bayes theorem for dealing with uncertainty and incompleteness; anticipated halting of a hypothesis investigation when a satisfaction point is reached in determining its plausibility; and the possibility of "volunteering" informations not asked by the system. The combined and intelligent use of all these characteristics may be quite advantageous.

SUMÁRIO

1.	Introdução.....	1
1.1	"Background" e evolução histórica.....	1
1.2	Objetivos e justificativas da dissertação.....	4
1.3	Esboço da dissertação.....	5
2.	Idéias chave de IntForBayes.....	6
2.1	Escolhas feitas e o "porquê" delas.....	6
2.2	Inferência bayesiana.....	10
2.3	Importância de um fato.....	17
2.4	Eliminando alguns mal-entendidos sobre os S.E.'s bayesianos.....	19
2.5	Características de IntForBayes.....	23
2.6	Quando escolher IntForBayes.....	24
3.	Manual do usuário.....	26
3.1	Introdução.....	26
3.2	Tutorial.....	29
3.2.1	Ativação do sistema.....	29
3.2.2	Sessão típica.....	29
3.2.2.1	Editando uma pequena base de conhecimento.....	30
3.2.2.2	Executando um pequeno S.E.....	46
3.3	Alguns conselhos úteis.....	61
4.	Projeto e implementação.....	63
4.1	Estruturas de dados.....	63
4.2	Algoritmos principais de IntForBayes.....	72
4.2.1	O nível superior de IntForBayes.....	72
4.2.2	Módulo de edição.....	72

4.2.3	Módulo de inferências.....	73
4.2.3.1	Módulo principal de Inferências.....	73
4.2.3.2	Processamento de uma consulta.....	74
4.2.3.3	Propagação do grau de certeza de um fato.....	76
4.2.3.4	Oferta/correção de dados (voluntariamente).....	77
4.2.3.5	Investigação de um fato.....	79
5.	Testes de desempenho (número de perguntas X precisão)....	82
5.1	Evolução histórica dentro do projeto IntForBayes....	82
5.2	Aspectos a serem avaliados no algoritmo.....	86
5.3	Duas bases de conhecimento usadas nos testes.....	87
5.4	Planejamento dos testes.....	89
5.5	Conclusões sobre os testes.....	91
6.	Conclusões e sugestões.....	97
6.1	Conclusões.....	97
6.2	Sugestões para trabalhos futuros.....	99
	Referências bibliográficas.....	103
Apêndice A	Constantes de configuração de IntForBayes.....	108
Apêndice B	Resultados dos testes.....	109

1. INTRODUÇÃO

1.1 "BACKGROUND" E EVOLUÇÃO HISTÓRICA

Dentre as diversas sub-áreas da Inteligência Artificial a que tem tido maior destaque e sucesso prático, sob o ponto de vista comercial, é aquela dos chamados Sistemas Especialistas. "Sistemas Especialistas (S.E.'s) são sistemas inteligentes de computação que, num estreito mas profundo e complexo domínio de aplicação, capturam o conhecimento de peritos (e como estes fazem uso dele), de modo que, ao serem usados, estes sistemas oferecem desempenho em tudo comparável aos daqueles peritos. Entre os principais tipos de S.E.'s podemos citar os diagnosticadores, os classificadores, os planejadores, os projetistas, os preditores, os monitoradores e os controladores."

O primeiro dos S.E.'s foi o DENDRAL [BUCH 78] (interpretador de espectrograma de massa) e o que teve maior influência e repercussão foi o MYCIN [SHOR 78] (diagnosticador das causas e prescritor dos medicamentos/tratamentos em casos de infecções bacterianas no sangue) [FORS 84]. Um outro sistema de sucesso foi o PROSPECTOR [GASC 81] (avaliador do potencial mineralógico de uma região), sendo o primeiro a incorporar a inferência bayesiana na propagação de incertezas.

Para acelerar e facilitar o desenvolvimento de S.E.'s muitas ferramentas têm sido desenvolvidas. Provavelmente as mais populares são os "shells", que são constituídas de um motor de

inferência e uma base de conhecimento vazia (a ser preenchida), juntamente com algumas facilidades de edição, depuração e explanação ("como chegou a este resultado?" e "para que esta pergunta?"). Alguns "shells" foram obtidos a partir de S.E.'s já existentes, como é o caso do EMYCIN ("Empty Mycin") [MELL 80] (que foi obtido do MYCIN) e do KAS ("Knowledge Acquisition System") [DUDA 79] (derivado do Prospector). Estas ferramentas possuem fortes vinculações com os sistemas que as originaram, tornando as suas utilizações mais restritivas. Outros "shells" foram desenvolvidos sem muita vinculação com outros sistemas existentes, como é o caso do VP-EXPERT [HARM 88]. Além dessa classe de ferramentas existem as linguagens, que possuem maior flexibilidade no desenvolvimento de S.E.'s e que podem ser mais voltadas para os S.E.'s (exemplos OPSS [FORG 80]), LIDIA [FERR 90] ou menos voltadas (linguagens de propósito geral, tais como Pascal, ou de Inteligência Artificial, tais como Lisp e Prolog). O uso de linguagens é mais complexo e exige maiores investimentos (em tempo, esforço e dinheiro) que o uso dos "shells". Uma avaliação cuidadosa deve ser feita para descobrir quais características das ferramentas são necessárias para o desenvolvimento de uma certa tarefa [ROTH 89].

Para que o conhecimento de um especialista possa ser capturado, conservado e manipulado, é conveniente representá-lo em uma forma que facilite estas tarefas. As representações mais utilizadas e que têm dado os melhores resultados são as seguintes:

- 1) Representação Lógica: expressa o conhecimento em algum tipo de

lógica (proposicional de primeira ordem, nebulosa, dos "defaults", etc); utiliza induções, deduções, etc para obter informações a partir dos dados conhecidos.

- 2) Representação Procedural: o conhecimento é formulado em termos de uma sequência de transformações codificadas em procedimentos (receitas) que dizem como o conhecimento deve ser manipulado.
- 3) Sistemas de Produção: o conhecimento é codificado com o uso de regras do tipo: SE uma condição é satisfeita ENTÃO uma ação é realizada ou uma conclusão é obtida.
- 4) Quadros ("Frames"): são estruturas de dados estáticas que contêm informações de como utilizar tais estruturas e quais as ações que devem ser desempenhadas quando certas condições são satisfeitas.
- 5) Redes Semânticas: a representação do conhecimento é feita através de grafos onde os nós representam objetos ou situações e os arcos indicam as relações entre eles. Nesta classificação estão situadas as chamadas Redes Bayesianas, que são grafos acíclicos direcionados e ponderados, nos quais os nós representam proposições (ou variáveis), os arcos representam a existência de influências diretas entre as proposições ligadas e os pesos dessas influências são quantificados pelas probabilidades condicionais [QUIN 82].

1.2 OBJETIVOS E JUSTIFICATIVAS DA DISSERTAÇÃO

Este trabalho tem como objetivos a concepção, a implementação e a validação de uma ferramenta (do tipo "shell") para desenvolvimento, manutenção e execução de S.E.'s bayesianos. Ela deve servir para testes e avaliação de várias idéias visando maior eficiência e facilidade na sua utilização tanto pelo engenheiro de conhecimento, como pelo especialista, como pelo usuário final. Desta forma pretende-se contribuir para o enriquecimento dos suportes atualmente disponíveis na área de Inteligência Artificial (I.A.) e S.E.'s.

A validade deste trabalho justifica-se: pelo seu caráter de pesquisa, constituindo-se num laboratório de testes de idéias; pelo seu alcance de uso em aplicações diferentes; fácil adaptação ao sistema, mesmo por pessoas sem muitos conhecimentos nesta área; facilidade e rapidez no desenvolvimento, manutenção e execução de S.E.'s bayesianos; simplicidade no ajuste do equilíbrio desejado entre precisão e tempo gasto com perguntas no processo de investigação de hipóteses. Neste processo é empregada uma seleção criteriosa na escolha dinâmica da ordem e do número de perguntas formuladas, procurando imitar o caráter inteligente e eficiente dos especialistas humanos.

O trabalho de [NUNE 86] motivou o Grupo de Inteligência Artificial do DSC/UFPB. Neste, foram desenvolvidos um S.E. bayesiano (SINDROMUS [NICO 87]) e uma linguagem para desenvolvimento de S.E.'s bayesianos (LIDIA [FERR 90]). Nosso trabalho é a complementação natural deles.

1.3 ESBOÇO DA DISSERTAÇÃO

No capítulo 2 (Idéias Chave de IntForBayes) discutimos as escolhas conceituais feitas para a ferramenta, além do porquê delas; introduzimos a inferência bayesiana e o conceito de importância de um fato; procuramos eliminar alguns mal-entendidos sobre S.E.'s bayesianos; listamos as características principais de IntForBayes; e definimos os domínios de conhecimento ideais para seu uso.

No capítulo 3 (Manual do Usuário) apresentamos, através de um detalhado exemplo, o funcionamento de todas as opções de todos os "menus" de IntForBayes, usadas tanto na edição de uma base de conhecimento como na consulta à mesma. Esboçamos também uma metodologia para o bom uso do "shell", na construção e consulta a bases de conhecimento.

No capítulo 4 (Projeto e Implementação) descrevemos, em alto nível, as estruturas de dados e os algoritmos principais do sistema.

No capítulo 5 (Testes de Desempenho) consideramos os aspectos a serem avaliados no algoritmo, descrevemos os testes feitos, discutimos os resultados e extraímos conclusões sobre os mesmos.

Finalmente, no capítulo 6 (Conclusões e Trabalhos Futuros), apresentamos as conclusões gerais sobre a ferramenta e listamos algumas sugestões para a continuação deste trabalho visando o seu enriquecimento.

que dizem respeito ao problema são incompletas e aproximadas [BHAT 90]. Algumas teorias utilizadas para manipular incompletudes e incertezas de informações ("raciocínio aproximado") são a Teoria da Probabilidade, a Teoria da Evidência de Shafer [SHAF 76] e a Teoria da Possibilidade de Zadeh [ZADE 78]. Na Teoria da Probabilidade a atualização das probabilidades das hipóteses é obtida a partir das suas probabilidades a priori e das probabilidades condicionais de sua evidências (ver 2.2). A Teoria da Evidência [SHAF 76] é uma extensão da Teoria da Probabilidade onde é feita uma distinção entre "ignorância" e "incerteza" e é introduzido o conceito de "funções de crença" e limites na atribuição das probabilidades dos eventos em vez de especificar as probabilidades exatamente. A Teoria da Possibilidade foi desenvolvida por Zadeh como uma aplicação da Teoria dos Conjuntos Nebulosos para representar incertezas em algum termo linguístico. Em nosso trabalho optamos por fazer uso da Teoria da Probabilidade pela sua simplicidade, forte embasamento matemático, apresentação de resultados satisfatórios e facilidade de implementação.

Há dois modos extremos de encadeamento do raciocínio:

- 1) no modo dirigido pelos objetivos, também conhecido como regressivo ou "backward chaining", solicitamos que o sistema investigue uma hipótese e ele, raciocinando na direção objetivo \rightarrow fatos (ou seja, perguntando-se "para provar esta hipótese, preciso provar o quê? E para provar isto, preciso provar o quê? ...") chega às perguntas influenciadoras daquela hipótese. No processo, é permitido ao sistema propagar os

graus de certeza somente nos arcos que fazem o caminho da pergunta atual à hipótese escolhida).

- 2) no modo dirigido pelos dados, também conhecido como progressivo ou "forward chaining", informamos ao sistema o que conhecemos sobre as evidências e ele, raciocinando na direção fatos → conclusões (ou seja, perguntando-se "quais são as conclusões diretas que posso tirar do que sei? e, daí, quais as conclusões que se seguem? ...") chega a todas as conclusões que podem ser inferidas dos fatos conhecidos. No processo, o sistema é exigido propagar os graus de certeza em todos os arcos de todos os caminhos ascendentes que se iniciem nas evidências conhecidas.

Ambos estes modos têm suas vantagens e são indicados para certas classes de aplicação. Visando uma ferramenta que combinasse as vantagens de ambos os modos, fizemos as escolhas de: (1) permitirmos o encadeamento "forward" na fase de "voluntariamento" (onde o usuário toma a iniciativa de fornecer espontaneamente dados não solicitados pelo sistema) ou de correções dos dados; (2) permitirmos a escolha de uma hipótese a ser investigada na agradável forma (típica do encadeamento "backward") que evidencia ter um objetivo; e (3) propagarmos os graus de certeza de cada nodo para todos os seus pais. Assim, embora o "For" do nome IntForBayes reflita somente a capacidade dele fazer inferências no modo "forward", na verdade ele poderá atuar no contínuo desde 100% "forward" (se voluntariarmos todas as informações e não escolhermos nenhuma hipótese) até 100% "backward" (se escolhermos uma hipótese e não voluntariarmos nada).

No processo de investigação de hipóteses, a escolha da ordem dos fatos analisados pode ser cega ou seguir certos critérios de modo a espelhar aqueles seguidos pelos especialistas na resolução de seus problemas. IntForBayes procura perseguir os fatos em uma ordem decrescente do "ganho de informação" que estes fatos podem proporcionar no objetivo de esclarecer a hipótese sendo investigada. A ordenação desses fatos é feita dinamicamente utilizando o valor heurístico "Importância do Fato", definido em 2.3 e justificado em 5.5.

Além da escolha da ordem dos fatos analisados, no processo de investigação de hipóteses precisa ser definida a quantidade de fatos a serem considerados nesta investigação. Ou optamos por considerar todos os fatos que influenciam a hipótese investigada, ou consideramos apenas alguns deles. Os bons especialistas procuram sempre chegar às suas conclusões com o menor número de perguntas possíveis (isto implica, na maioria dos casos, em economia de tempo e dinheiro, menor incômodo, maior produtividade, etc.). Pensando nisso, IntForBayes faz uso de um ponto de satisfação (facilmente ajustável pelo usuário segundo suas restrições de números de perguntas ou precisão a ser obtida) que pode (quando convenientemente ajustado) reduzir sensivelmente o número de perguntas formuladas no processo de investigação de hipóteses, sem comprometer significativamente os resultados obtidos, uma vez que as perguntas que mais contribuem neste objetivo são feitas logo no início desse processo.

Os S.E.'s podem não permitir voluntariamente, ou permiti-lo em alguns poucos pontos fixos de uma consulta, ou ainda permiti-

lo sempre que for desejado. Em IntForBayes é dado um destaque especial ao voluntariamento, de modo que tanto é possível como aconselhado que ele seja invocado sempre que o usuário deseje oferecer e/ou corrigir informações. Este destaque deve-se ao fato de que uma informação prestada corretamente e tão logo quanto possível afunilará rapidamente o universo de possíveis candidatas à solução (ou soluções), proporcionando uma convergência mais rápida na parada do processo de investigação.

2.2 INFERENCIA BAYESIANA

Apresentamos aqui fundamentos e resultados teóricos que nortearam a concepção da ferramenta.

Dada uma hipótese H, define-se sua chance C(H) pela razão entre a probabilidade a favor e a probabilidade contra a hipótese:

$$C(H) = \frac{P(H)}{1 - P(H)} \quad (1)$$

Chamamos de chance a priori de uma hipótese H, $C_p(H)$, a chance dela ser verdadeira quando não se conhece qualquer evidência a favor ou contra a mesma. Por exemplo, se, ao tomarmos uma pessoa ao acaso dentre aquelas que vêm a um posto de saúde, a probabilidade dela estar com sarampo é 3,0%, então:

$$C_p(\text{sarampo}) = 0,03 / (1 - 0,03) = 0,031$$

O uso de chance em lugar de probabilidade simplifica as equações a que chegaremos, trazendo vantagens computacionais.

Mostramos, a seguir, em quatro etapas de generalização e com direcionamento para a explicação da ferramenta, como se procede a inferência bayesiana. Estas idéias podem ser encontradas na literatura (vide, por exemplo, [NUNE 86], [DUDA 79]). No entanto, na última etapa, nós formulamos as expressões de modo a obter maior eficiência na execução dos cálculos.

Primeiramente, o teorema de Bayes permite-nos calcular a chance (a posteriori) de uma hipótese H, conhecida a verdade ou a falsidade de uma evidência E:

$$C(H:E) = C_p(H) * FS(E,H) \quad (*) \quad (2.a)$$

$$C(H:\sim E) = C_p(H) * FN(E,H) \quad (2.b)$$

onde FS (Fator do Sim ou Suficiência Lógica) e FN (Fator do Não ou Necessidade Lógica) são dados por:

$$FS(E,H) = \frac{P(E:H)}{P(E:\sim H)} \quad (3.a)$$

$$FN(E,H) = \frac{P(\sim E:H)}{P(\sim E:\sim H)} = \frac{1 - P(E:H)}{1 - P(E:\sim H)} \quad (3.b)$$

Em segundo lugar, suponhamos que, ao invés de sabermos necessariamente que uma evidência E é verdadeira ou falsa, ela seja conhecida com um grau de certeza, GC(E), no intervalo [-1,+1] (este intervalo foi escolhido de modo a simplificar algumas das equações a que chegaremos). Fazendo-se uma interpolação linear tal que os valores -1, 0 e +1 de GC sejam mapeados nos multiplicadores FN, 1 e FS, respectivamente,

 (*) Lê-se C(H:E) como "chance de H dado E".

obtemos:

$$C(H:E) = C_p(H) * \text{FatorAtualizador}(E,H) \quad (**) \quad (4)$$

onde:

$$- \text{FatorAtualizador}(E,H) = \begin{cases} (FS(E,H) - 1) * GC(E) + 1 & \text{se } GC(E) \geq 0 \\ (1 - FN(E,H)) * GC(E) + 1 & \text{se } GC(E) < 0 \end{cases} \quad \begin{matrix} (5.a) \\ (5.b) \end{matrix}$$

Exemplo: Supondo que Febre Alta é conhecida com $GC = 0,80$ e que

$$P(\text{Febre Alta} : \text{Sarampo}) = 0,40$$

$$P(\text{Febre Alta} : \sim\text{Sarampo}) = 0,04$$

então podemos calcular a chance a posteriori de Sarampo dado $GC(\text{Febre Alta}) = 0,80$:

$$FS(\text{Febre Alta}, \text{Sarampo}) = 0,40 / 0,04 = 10,0$$

$$\text{FatorAtualizador}(\text{Febre Alta}, \text{Sarampo}) = (10,0 - 1) * 0,80 + 1 = 8,2$$

$$C(\text{Sarampo} : \text{Febre Alta}) = 0,031 * 8,2 = 0,254$$

(Esta chance corresponde à probabilidade 20,26%).

Notemos, em um terceiro passo, que uma vez calculada a chance de H conhecidas m de suas evidências, ou seja, $C(H:E_1, E_2, \dots, E_m)$, $m \geq 1$, e uma vez conhecido o GC de uma nova evidência E_{m+1} , podemos atualizar a chance de H calculando $C(H:E_1, E_2, \dots, E_{m+1})$ pela expressão (4), onde $C_p(H)$ passa a ser a chance atual de H, $C_a(H) = C(H:E_1, E_2, \dots, E_m)$. Pode-se, portanto, escrever:

$$C(H:E_1, E_2, \dots, E_{m+1}) = C_p(H) * \prod_{i=1}^{m+1} \text{FatorAtualizador}(E_i, H) \quad (6)$$

(**) Doravante, entender-se-á $C(H:E)$ como "chance de H dado que E é conhecida com um certo grau de certeza".

ou equivalentemente:

$$C(H:E1, \dots, E_{m+1}) = Ca(H) * \text{FatorAtualizador}(E_{m+1}, H) \quad (7)$$

onde $Ca(H)$ é a chance de H anterior ao conhecimento da $(m+1)$ -ésima evidência.

Consideremos finalmente uma rede bayesiana genérica, comportando hipóteses em vários níveis (uma rede particular é mostrada na figura 2).

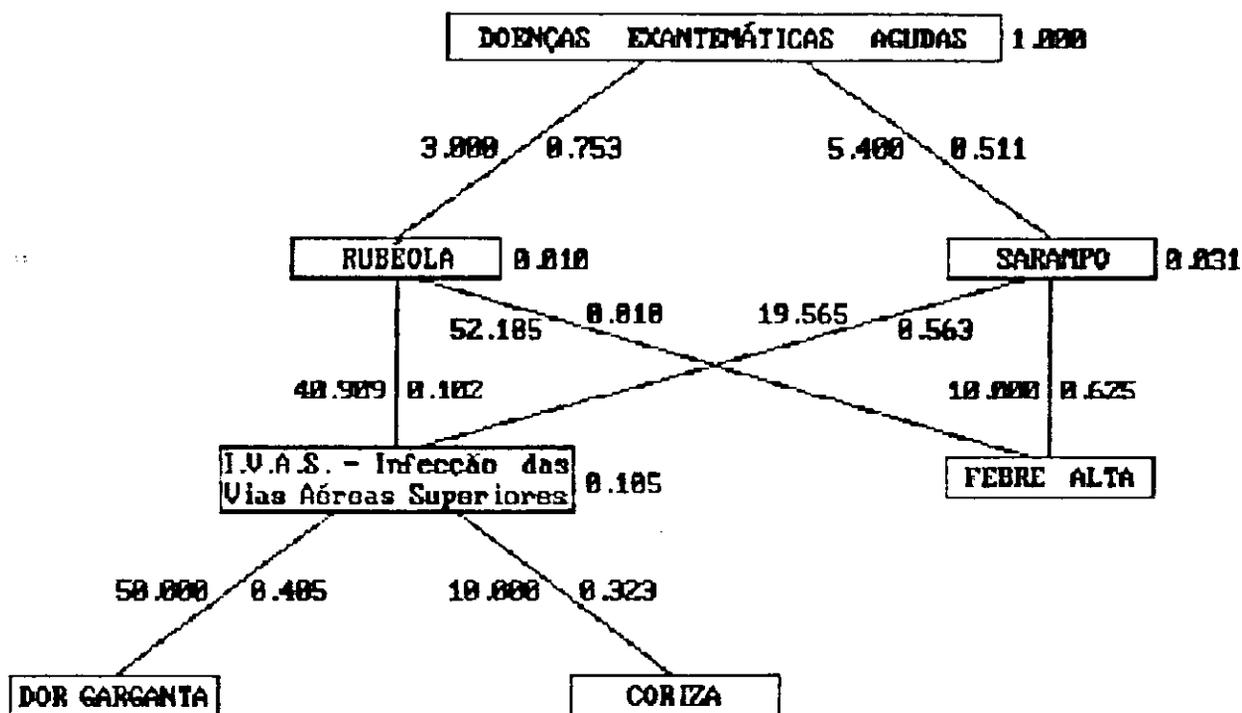


FIGURA 2 - Modelo de uma Rede Bayesiana

A terminologia utilizada para descrever os diversos tipos de nodos de uma rede de conhecimento é a seguinte:

- 1) Evidências ou fatos perguntáveis: são os nodos terminais da rede de conhecimento e cujos graus de certeza podem ser diretamente informados pelo usuário ("DOR GARGANTA", "CORIZA", e "FEBRE ALTA").

- 2) Hipóteses ou fatos dedutíveis: são os demais nodos da rede de conhecimento e cujos graus de certeza são obtidos a partir das evidências ou de outras hipóteses ("EXANTEMAS AGUDOS", "RUBEDOLA", "SARAMPO", e "I.V.A.S.").
- 3) Raiz absoluta: é o nodo particular que não tem pai. Seu nome é o mesmo dado ao S.E. ("DOENÇAS EXANTEMATICAS AGUDAS").
- 4) Hipóteses de primeiro nível: são as hipóteses que podem ser selecionadas pelo usuário para alguma investigação. Estas hipóteses correspondem aos nodos filhos da raiz absoluta ("RUBEDOLA" e "SARAMPO").
- 5) Hipóteses intermediárias: são hipóteses filhas das hipóteses de primeiro nível ou de outras hipóteses intermediárias ("I.V.A.S.").

Além das evidências, aos demais fatos (hipóteses) devem ser também atribuídos graus de certeza para que possamos utilizar, com fins de generalização, as mesmas fórmulas apresentadas; isto é feito por interpolação linear tal que as probabilidades 0, $P_p(F)$ (probabilidade a priori do fato F) e +1 sejam mapeadas respectivamente nos graus de certeza -1, 0 e +1:

$$GC(F) = \begin{cases} \frac{Ca(F) - Cp(F)}{1 + Ca(F)} & \text{se } Ca(F) \geq Cp(F) & (8.a) \\ \frac{Ca(F) - Cp(F)}{Cp(F) * (1+Ca(F))} & \text{se } Ca(F) < Cp(F) & (8.b) \end{cases}$$

Ora, sendo mais vantajoso fazer as perguntas sobre as evidências em uma ordem inteligente e numa quantidade parcimoniosa do que fazê-las em uma ordem cega e numa quantidade exaustiva, introduziremos em 2.3 o conceito de importância de um

fato, recalculada a cada nova informação prestada sobre uma das suas evidências. Como a fórmula (6) não é muito eficiente para estes recálculos, devemos usar a (7). No entanto, como ela não é muito eficiente para modificações nas evidências, propomos inicializar as chances de todas as hipóteses com suas chances a priori e a cada atualização da informação sobre um fato F (o que resulta de uma informação prestada sobre uma evidência), cada hipótese H que depende diretamente de F terá sua chance, C(H), atualizada por

$$C(H) := C(H) * \frac{FAuxChanceNovo(F,H)}{FAuxChanceVelho(F,H)} \quad (9)$$

onde:

$$- FAuxChanceNovo(F,H) = \begin{cases} (FS(F,H) - 1) * GC(F) + 1 & \text{se } GC(F) \geq 0 \\ (1 - FN(F,H)) * GC(F) + 1 & \text{se } GC(F) < 0 \end{cases} \quad \begin{matrix} (10.a) \\ (10.b) \end{matrix}$$

- FAuxChanceVelho(F,H) é inicializado com o valor 1 e assume sucessivamente os valores imediatamente anteriores de FAuxChanceNovo(F,H).

Note-se que (9) é uma atribuição computacional e não uma equação matemática, é eficiente e tem uma vantajosa propriedade: se o usuário desejar corrigir uma informação já prestada, modificando o grau de certeza de uma evidência, tudo se passará de uma maneira muito natural, como se ele o estivesse anunciando pela primeira vez; os efeitos da informação anterior sobre esta evidência serão automaticamente eliminados.

Consideremos, para fins ilustrativos, a rede da figura 2. Os valores mostrados ao lado de cada nodo correspondem às chances a priori e os pares mostrados em cada arco correspondem ao par (Fator do Sim, Fator do Não) do arco. Suponhamos que seja informada a evidência DDR GARGANTA com grau de certeza 0.8. Esta informação propaga-se na seguinte sequência:

$$FAuxChanceNovo(DDR\ GARGANTA, IVAS) = (50 - 1) * 0,8 + 1 = 40,2$$

$$C(IVAS) = \frac{0,105 * 40,2}{1} = 4,221$$

$$GC(IVAS) = \frac{4,221 - 0,105}{1 + 4,221} = 0,788$$

$$FAuxChanceNovo(IVAS, SARAMPO) = (19,565 - 1) * 0,788 + 1 = 15,629$$

$$C(SARAMPO) = \frac{0,031 * 15,629}{1} = 0,485$$

$$GC(SARAMPO) = \frac{0,485 - 0,031}{1 + 0,485} = 0,305$$

$$FAuxChanceNovo(IVAS, RUBEDLA) = (40,909 - 1) * 0,788 + 1 = 32,448$$

$$C(RUBEDLA) = \frac{0,01 * 32,448}{1} = 0,324$$

$$GC(RUBEDLA) = \frac{0,324 - 0,010}{1 + 0,324} = 0,237$$

...

Se agora fosse informada a evidência CORIZA com grau de certeza 0,2, obteríamos:

$$FAuxChanceNovo(CORIZA, IVAS) = (10 - 1) * 0,2 + 1 = 2,8$$

$$C(IVAS) = \frac{4,221 * 2,8}{1} = 11,819$$

$$GC(IVAS) = \frac{11,819 - 0,105}{1 + 11,819} = 0,914$$

$$FAuxChanceNovo(IVAS, SARAMPO) = (19,565 - 1) * 0,914 + 1 = 17,968$$

$$C(SARAMPO) = \frac{0,485 * 17,968}{15,629} = 0,558$$

$$GC(SARAMPO) = \frac{0,557 - 0,031}{1 + 0,557} = 0,338$$

...

2.3 IMPORTANCIA DE UM FATO

Quando um especialista humano - um médico, por exemplo - quer fazer um diagnóstico com rapidez e eficiência, ele procura pesquisar o menor número possível de fatos; para isto, faz-se necessária uma inteligente sequência de indagações. Em geral, pode-se afirmar que "melhor o perito menor o número de evidências de que ele precisa para chegar a uma conclusão".

Para imitar um bom perito humano, um S.E. deve ter também essa eficiência quanto ao número e à ordem de perguntas que ele precisa fazer ao usuário.

Ao tentar provar uma hipótese, IntForBayes necessita de um critério para definir qual dos fatos diretamente relacionadas com aquela hipótese deve ser investigado primeiro. Isto se faz atribuindo-se a cada fato um valor heurístico local chamado importância do fato em relação à hipótese:

$$Imp(F,H) = (GC(H:F.Teto) - GC(H:F.Piso))/2 \quad (11)$$

Escolhemos esta fórmula, entre diversas outras opções, após a constatação do seu melhor desempenho obtido nos testes. (5.1 relata a evolução e a razão que levou a esta fórmula. 5.5 demonstra sua vantagem). Traduzindo em palavras a equação acima, a importância de F para H é a diferença, mapeada no intervalo [0,+1], entre o grau de certeza de H dado que E atinja seu Teto e o grau de certeza de H dado que F atinja seu Piso, sendo Teto e Piso de um fato os valores máximo e mínimo que sua chance ainda poderá alcançar, respectivamente. Teto e Piso são inicializados conforme as definições acima, mas são atualizados por fórmulas bem mais eficientes, pois em cada atualização aproveitam os cálculos feitos anteriormente. Teto é atualizado segundo a fórmula:

$$\text{Teto}(H) := \text{Teto}(H) * \frac{\text{FRuxTetoNovo}(F,H)}{\text{FRuxTetoVelho}(F,H)} \quad (12)$$

onde:

$$\text{FRuxTetoNovo}(F,h) = \begin{cases} \text{FRux1}(F,N) & \text{se } \text{FRux1} > \text{FRux2} & (13.a) \\ \text{FRux2}(F,N) & \text{caso contrário} & (13.b) \end{cases}$$

$$\text{FRux1}(F,H) = \begin{cases} (\text{FS}(F,H) - 1) * \text{GC}(F) + 1 & \text{se } F \text{ é um fato perguntável} & (14.a) \\ (\text{FS}(F,H) - 1) * \text{GC}(F.\text{Teto}) + 1 & \text{se } F \text{ é um fato dedutível} & (14.b) \end{cases}$$

(onde GC(F.Teto) é o GC que F teria se alcançasse seu Teto)

$$\text{FAux2}(F, H) = \begin{cases} (\text{FN}(F, H) - 1) * \text{GC}(F) + 1 & \text{se } F \text{ é um fato perguntável} & (15.a) \\ (\text{FN}(F, H) - 1) * \text{GC}(F.\text{Piso}) + 1 & \text{se } F \text{ é um fato dedutível} & (15.b) \end{cases}$$

(onde $\text{GC}(F.\text{Piso})$ é o GC que F teria se alcançasse seu Piso)

$\text{FAuxTetoVelho}(F, H)$ é inicializado, no caso de F ser folha da árvore, com $\text{FS}(F, H)$ (se $\text{FS}(F, H)$ for maior que $\text{FN}(F, H)$), e com $\text{FN}(F, H)$ (caso contrário). No caso de F não ser folha é inicializado com $\text{FAuxTetoNovo}(F, H)$ e depois assumirá sucessivamente os valores anteriores de $\text{FAuxTetoNovo}(F, H)$.

Piso é atualizado similarmente.

2.4 ELIMINANDO ALGUNS MAL-ENTENDIDOS SOBRE OS S.E.'s BAYESIANOS:

A seguir, responderemos às críticas mais frequentemente feitas aos S.E.'s bayesianos, resultantes de alguns mal-entendidos. Enfatizamos que nosso objetivo não é atacar outros S.E.'s, mas apenas defender um espaço para os S.E.'s bayesianos.

Pergunta: Regras SE ... ENTÃO ... são muito mais comuns e populares para S.E.'s. Isto não reflete o fato que S.E.'s bayesianos são sempre piores, em algum sentido?

Resposta: Não. Cada modo de representar e fazer inferências sobre o conhecimento tem seu lugar, ao menos para certas classes de domínios de conhecimento. Na verdade regras são muito mais

naturais para muitas áreas de conhecimento sem incerteza/incompletude, foram consagradas nestas aplicações e daí (e também por razões históricas) dominaram a cena também para conhecimento com incerteza/incompletude. Mas, pelo menos em muitos destes últimos domínios, redes bayesianas não são em nada inferiores às regras:

- a) Têm um sólido e respeitável embasamento matemático (na teoria das probabilidades);
- b) Têm precisão, apresentam resultados satisfatórios aos especialistas;
- c) Não são menos naturais/fáceis de usar. Aliás, o conceito de naturalidade depende muito do perfil do especialista e de treinamento, e em certos domínios temos achado mais natural perguntar ao especialista:

" - Qual a prevalência (mesmo que estimada) de sarampo nas crianças da região? 1:1.000?

- Qual a porcentagem de sarampo entre as crianças que apresentam ...? 1:2 ?"

do que perguntar-lhe:

" - Qual o fator de atenuação (F.A.) que devo dar à regra

SE criança apresenta ... ENTÃO sarampo com F.A. ...

isto é, supondo que você creia com 100% de certeza que uma criança apresenta ..., com que grau de certeza você creia que ela tem sarampo ?"

- d) Sistemas bayesianos combatem duas grandes e más tendências presentes na construção (especialmente por novatos) de S.E.'s baseados em regras: Escrever uma regra

SE Cond1 & ... & Cond6 ENTÃO Concl COM FA = 98%

e esquecer de escrever:

- 1) - algumas das regras necessárias e relevantes entre as (2 elevado a 6) - 1 = 63 potenciais regras envolvendo os subconjuntos das condições, por exemplo:

SE Cond1 ENTÃO Concl com FA = 50%

...

SE Cond2 & ... & Cond6 ENTÃO Concl COM FA = 72%

- 2) - as regras de negação (tão importantes quanto as regras de afirmação):

SE NÃO (Cond1) ENTÃO Concl COM FA = - 13%

...

SE NÃO (Cond1 & ... & Cond6) ENTÃO Concl COM FA = - 28%

Pergunta : O teorema de Bayes assume que os filhos de cada fato sejam rigorosamente independentes. Como no mundo real isto muitas vezes não ocorre tão rigorosamente, os sistemas bayesianos não se aplicam somente a áreas irrealis, sem interesse prático?

RESPOSTA : Não. Aliás críticas igualmente severas poderiam ser feitas aos mecanismos de inferência de todas as outras representações de conhecimento. Por exemplo: "Os sistemas baseados em regras com inferência a la MYCIN assumem que os graus de certeza das conclusões sejam rigorosamente iguais aos calculados por certas fórmulas. Como no mundo real isto muitas vezes não ocorre tão rigorosamente, ...". O ponto importante a notar é que na falta de modelos matemáticos que reflitam perfeitamente a realidade, podemos e devemos usar modelos aproximados que dêem resultados satisfatórios. E isto os sistemas

bayesianos (mesmo com simplificações), como os outros, podem fazer em muitas áreas reais e de interesse prático. Observamos ainda que:

- a) A maioria dos casos de grande não independência entre os filhos pode ser evitada com um certo cuidado na escolha do cardápio dos fatos: evitando os sinônimos (ou quase sinônimos) - se incluirmos Febre, não devemos incluir Temperatura Alta - e os antônimos (ou quase antônimos) - se incluirmos Sexo Masculino não devemos incluir Sexo Feminino.
- b) Medianos desvios de independência rigorosa resultam, após muitas perguntas, em pequenos desvios nos resultados, sem importância prática. Assim, embora Coriza e Dor de Garganta não sejam rigorosamente independentes, um sistema que levasse estas correlações em conta, após 10 perguntas poderia diagnosticar Gripe com grau de certeza 64%, enquanto um sistema simplificado diagnosticaria Gripe com grau de certeza 59%.
- c) Se desejarmos, os desvios da independência ideal podem ser aproximados pelo uso do nodo "operador de inferência" (que pode simular diversos graus de AND, OR, XOR, etc), já previsto para uma expansão do atual motor de inferência de IntForBayes. Ver a descrição de LIDIA, em [FERR 90].
- d) Se desejarmos, podemos introduzir um tratamento dos correlacionamentos estatísticos (ver [MACH 85], [NORU 75a] e [NORU 75b]).

Pergunta: A teoria de Bayes supõe que sabemos os valores das probabilidades dos eventos que muitas vezes não conhecemos. Isto

não inviabiliza o seu uso?

RESPOSTA: Não. Se tivermos os valores exatos dessas probabilidades, usemo-los; senão, usemos as melhores possíveis aproximações ou meros palpites, como fazemos com outras representações de conhecimento. Notemos que os resultados não são muito sensíveis a essas probabilidades. Por exemplo, se tomarmos as chances $C(H)$, $C(E14:H)$ e $C(E14:\sim H)$ dez vezes maiores que o devido, após coletarmos 10 a 14 evidências o grau de certeza de H pode estar só 1 a 5 % diferente do devido. Aliás, isto se repete com todos os mecanismos de inferência mais usados: o pior deles ainda dá ótimos resultados quando há bastante informações sobre a consulta e o conhecimento do especialista foi extraído com bastante profundidade.

2.5 CARACTERÍSTICAS DE INTFORBAYES

A seguir destacaremos algumas das principais características de IntForbayes:

- A informação voluntária de evidências é facultada ao usuário em vários pontos de uma consulta.

- O status de uma consulta pode ser guardado com vistas a uma posterior continuação da mesma.

- Informações prestadas podem ser corrigidas de um modo bastante natural e eficiente: com o algoritmo e as fórmulas utilizadas, a propagação de um novo grau de certeza de uma evidência anula automaticamente, em toda a rede, as influências

de informações anteriores sobre a mesma evidência.

- O sistema pode investigar sucessivamente qualquer uma das hipóteses, escolhida pelo usuário, a condição "default" correspondendo sempre à hipótese com maior grau de certeza.

- A investigação de uma hipótese só continua enquanto sua importância for maior que um certo limiar (ponto de satisfação), o qual é fixado em cada construção de um S.E. particular. Durante a consulta este limiar pode ser modificado. Sua diminuição leva a resultados mais precisos, a custo de maior número de perguntas.

- A implementação da "Investigação de uma Hipótese" identifica-se basicamente com o método de pesquisa conhecido como "Informed Depth First" (ver [NILS 71] e [KVIT 88]). A heurística é local e baseia-se na Importância definida em 2.3.

- A ordem dos fatos selecionados no processo de investigação de hipóteses é definida dinamicamente de modo que, a cada passo, seja escolhido aquele fato que proporcione o maior ganho esperado de informação.

2.6 QUANDO ESCOLHER INTFORBAYES

É consensual, entre os pesquisadores de Inteligência Artificial, que nenhum "shell" é ideal para todas as aplicações, e que para todo "shell" há aplicações onde ele é ideal. Os domínios de conhecimento ideais para IntForBayes, no seu estado atual, são aqueles que:

- 1) Têm objetivos de diagnosticar ou classificar;
- 2) Envolvem incerteza e incompletude, com pouco correlacionamento

entre as evidências de cada hipótese;

3) São de ordem 0 (isto é, não envolvem variáveis, nem expressões, nem funções/procedimentos).

Muitas dessas restrições poderão ser removidas futuramente (ver 6.2).

3. MANUAL DO USUÁRIO

Nesta seção é mostrado, de uma forma acessível mesmo a pessoas sem muito conhecimento na área de Sistemas Especialistas e/ou Computação, como utilizar IntForBayes na criação e uso de Sistemas Especialistas (S.E.'s).

3.1 INTRODUÇÃO

IntForBayes é um "shell" para desenvolvimento e execução de S.E.'s com inferência bayesiana, encadeamento "forward", inteligência na determinação da ordem e número das perguntas, e com a possibilidade de voluntariamento de evidências a qualquer ponto da consulta. Expliquemos melhor:

"Shell": um "shell" é uma ferramenta constituída de um motor de inferência e uma base de conhecimento vazia, juntamente com algumas facilidades de edição, depuração e explanação.

Inferência bayesiana: inferência é o processo de chegar a conclusões lógicas a partir de premissas ou fatos conhecidos. Na inferência bayesiana existe um formalismo para o raciocínio sobre conhecimento incompleto e com incerteza. Neste formalismo as proposições são quantificadas com parâmetros numéricos, significando o grau de crença à luz da parte do conhecimento capturada do especialista, e estes parâmetros são combinados e manipulados de acordo com as regras da teoria da probabilidade (ver 2.2).

Encadeamento "forward": no encadeamento "forward" o raciocínio caminha das evidências para as hipóteses: a cada novo fato conhecido fazem-se inferências sobre todas as hipóteses que com ele se relacionam direta ou indiretamente. Ou seja, tentamos concluir tudo que é possível a partir dos dados conhecidos (ver 2.1).

Inteligência na ordem das perguntas: uma das características dos especialistas na busca de soluções para um problema é em geral tentar obter primeiro as informações que possam contribuir da maneira mais decisiva ou substancial para a confirmação ou refutação de sua(s) suspeita(s). A ordem de perguntas feitas por IntForBayes é estabelecida em função de um valor heurístico, denominado de importância do fato, que procura capturar o potencial que fatos influenciadores de uma hipótese de primeiro nível escolhida possuem, para provocar alterações na hipótese em análise. Heurísticas são regras práticas ("macetes") que quase sempre funcionam aceitavelmente e tornam exequíveis as soluções que teórica e exatamente seriam impraticáveis.

Inteligência no número de perguntas: Segundo Kassier e Gorry [KASS 78] o número de fatos envolvidos no processo de raciocínio de um especialista é bem menor do que os envolvidos pelo não especialista. Um especialista chega a resultados muito bons fazendo poucas (além de inteligentes) perguntas e esta é uma das características mais desejadas nos sistemas especialistas. IntForBayes faz uso de um ponto de satisfação que, convenientemente escolhido, reduz sensivelmente o número de perguntas necessárias para se chegar a um resultado praticamente

tão confiável quanto aquele que seria obtido por exaustão, tornando o processo global de uma consulta muito mais eficiente.

Voluntariamento de evidências: o número de hipóteses que se destacam como mais promissoras para solução de problemas específicos numa determinada área de conhecimento, geralmente se torna, logo com as primeiras informações, muito pequeno se comparado com o universo de possíveis soluções. Por exemplo, em Medicina, o número normal de hipóteses promissoras após essas informações preliminares é da ordem de quatro ou cinco, e raramente ultrapassa seis ou sete [ELST 79]. A oferta voluntária de informações é chamada de voluntariamento. Se essas informações forem aquelas que nos saltam à vista (ou forem obtidas quase com custo zero) e, principalmente, se forem dadas logo no início da consulta, levarão a uma rápida convergência para o resultado, com um extraordinário ganho (em tempo / incômodo / dinheiro /...) em relação a uma busca exaustiva (em SINDROMUS [NICO 87], o voluntariamento de 2 a 3 informações permitiu geralmente a formação de boas suspeitas, confirmadas ou desconfirmadas com mais umas 2 a 10 perguntas, enquanto a busca exaustiva faria 700 perguntas!). IntForBayes permite que o usuário voluntarie informações em qualquer ordem e a qualquer momento. Este voluntariamento é não só possível como altamente aconselhado.

3.2 TUTORIAL

3.2.1 Ativação do Sistema

IntForBayes pode ser executado em microcomputadores compatíveis com o IBM-PC, sob sistema operacional compatível com DOS, com no mínimo 512 Kbytes de RAM e 1 disco flexível de 360 Kbytes. Aconselhamos usar pelo menos um AT, com monitor colorido.

IntForBayes é constituído de três módulos básicos denominados de IFB.EXE (módulo principal), EDITAR.EXE (módulo de edição) e INFERIR.EXE (módulo de inferências). Estes três módulos devem estar em um mesmo diretório da unidade (disco rígido, de preferência) em que se deseja trabalhar.

Para dar início a uma sessão de IntForBayes o usuário deve digitar IFB seguido de <Entra>. A partir de então o sistema entrará no ar e estará pronto para iniciar uma sessão.

3.2.2 Sessão Típica

Após a inicialização do sistema, conforme descrita na seção anterior, IntForBayes apresentará a seguinte tela:

IntForBayes - Ferramenta p/ construcao e execucao de SE's Bayesianos
Versao 1.0 - 1991 - UFPb/CCT/DSC

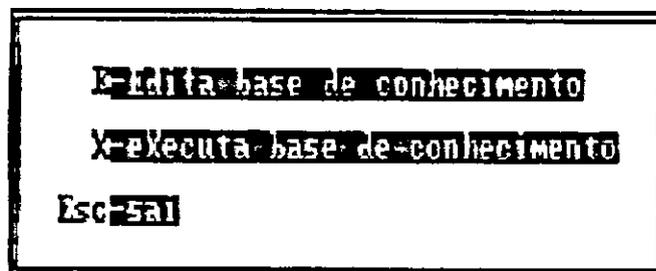


FIGURA 3.1 - "Menu" Principal de IntForBayes

Caso exista o desejo de criar uma nova base de conhecimento ou fazer manipulações (alterações, adições, eliminações, etc) como conhecimento de alguma base já criada por IntForBayes, o usuário deve utilizar nesta ocasião a opção de edição (E).

Caso queira utilizar o conhecimento de uma determinada base (previamente criada) para fazer algumas consultas, o usuário deve escolher a opção de execução (X).

3.2.2.1 Editando uma pequena Base de Conhecimento

Para formalizar o conhecimento de uma determinada área em uma forma que possa posteriormente ser usada, IntForBayes se utiliza de um editor específico que pode ser invocado pela opção de edição (E), após a sua inicialização.

Após a chamada do editor, IntForBayes apresentará a seguinte tela:

FASE DE EDIÇÃO

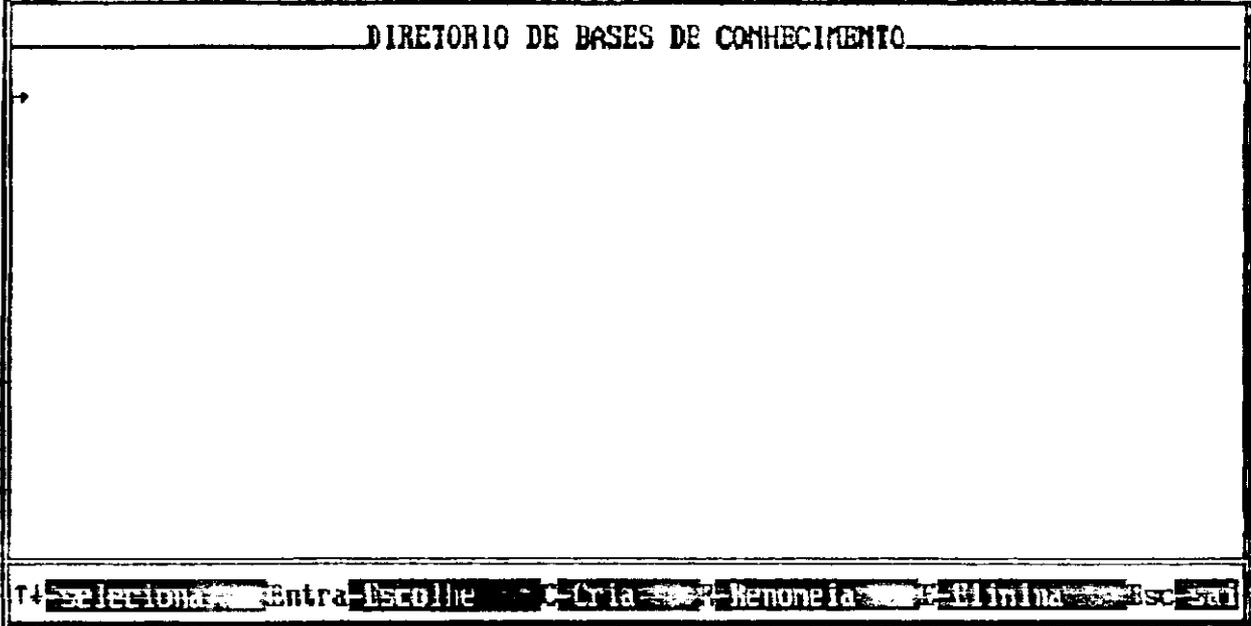


FIGURA 3.2 - "Menu" de Bases Existentes

Caso existam bases de conhecimento criadas anteriormente, os seus nomes serão apresentados na tela acima e poderá ser escolhida qualquer uma delas para receber alguma atualização, além de renomeação ou eliminação.

Nesta seção mostraremos, com um pequeno exemplo, como poderá ser criada uma base de conhecimento e, na seção seguinte, como utilizar esta base para fazer inferências com o conhecimento armazenado.

Suponhamos, a título de exemplificação, que queremos criar uma base de conhecimento que trate de doenças exantemáticas agudas. Doença exantemática aguda é qualquer doença aguda (início relativamente súbito) cujo sintoma mais nítido é o exantema (vermelhidão da pele). Exemplos: sarampo, rubéola e varicela.

A estrutura básica deste conhecimento é apresentada na seguinte figura:

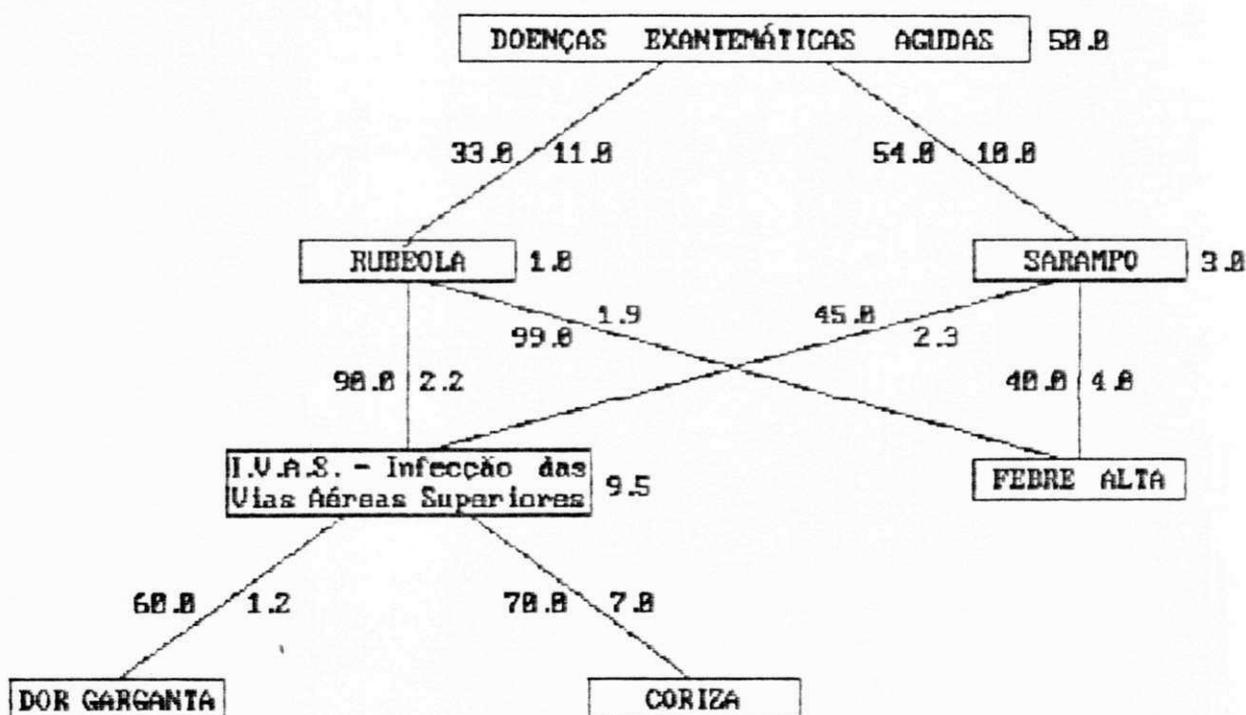


FIGURA 3.3 - Rede de Conhecimento de Doenças Exantemáticas Agudas

Na rede da figura acima cada nodo corresponde a um fato (perguntável ou dedutível) e cada arco corresponde a uma ligação de dependência entre dois fatos interligados. Os valores apresentados ao lado dos fatos são as probabilidades a priori dos fatos e os apresentados em cada arco (ligação) são as probabilidades condicionais de cada ligação. (IMPORTANTE: todas as probabilidades foram escolhidas com fins didáticos de mera ilustração de diferentes situações no uso do "shell", não tendo nenhum intuito de refletir a realidade e o conhecimento dos especialistas). Por exemplo, a probabilidade a priori de SARAMPO é igual a 3.0%; a probabilidade de alguém estar com FEBRE ALTA sabendo-se que tem SARAMPO é 40.0% e a probabilidade de alguém

estar com FEBRE ALTA sabendo-se que não tem SARAMPO é 4.0%

Utilizando-se a opção "C" (Cria base de conhecimento) da figura 3.2, o sistema irá solicitar o nome da nova base (são considerados no máximo 30 caracteres) que se deseja criar. Nesta ocasião digitamos "Doenças Exantemáticas Agudas".

Em seguida, escolhemos a base de conhecimento (recém criada) para a edição propriamente dita do conhecimento. Após esta escolha IntForBayes apresentará a seguinte tela:

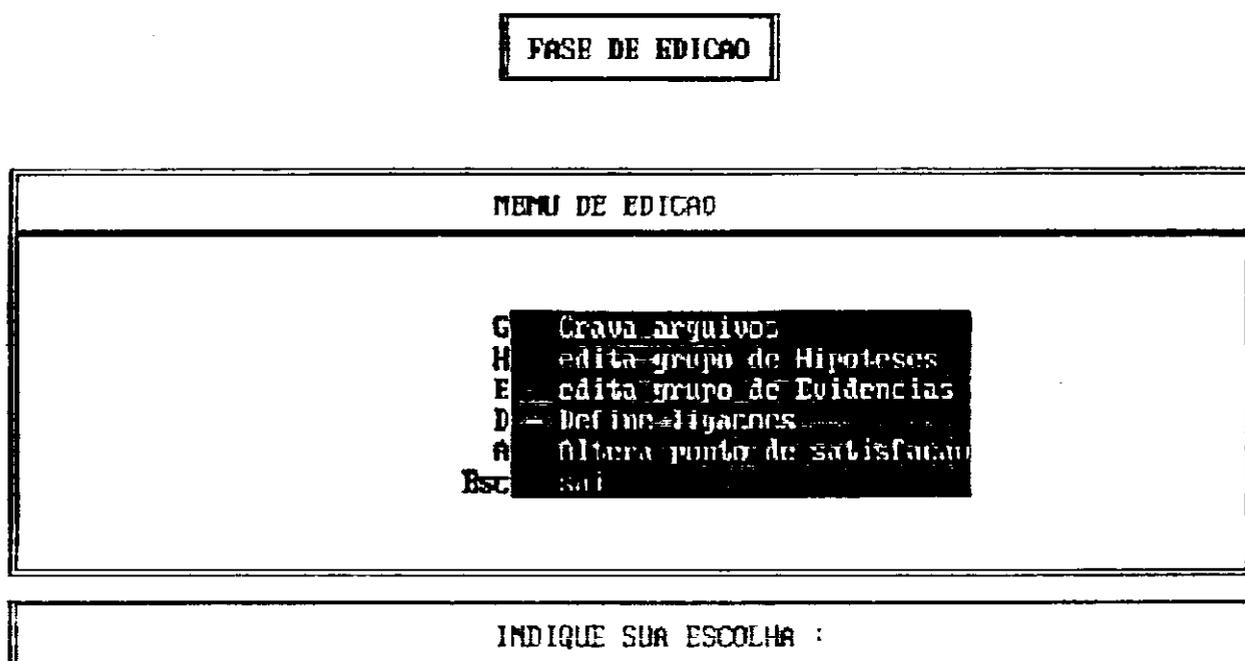


FIGURA 3.4 - "Menu" Principal de Edição

A opção "G" (Grava arquivos) deve ser usada para ratificar as alterações feitas durante a edição atual do conhecimento, gravando-as sobre a versão anterior, que será perdida. A saída do editor sem este comando implicará na manutenção da base de conhecimento anteriormente gravada, ou seja, as alterações

processadas após a última gravação serão ignoradas.

A quantidade de fatos (hipóteses ou evidências) que constituem uma base de conhecimento nos sistemas reais é em geral muito grande, na ordem de muitas centenas ou de alguns milhares. Para uma melhor estruturação destes fatos (e, assim, para uma maior facilidade de uso posterior), IntForBayes permite agrupá-los de acordo com a conveniência do usuário.

Usando-se a opção "H" (edita grupo de Hipóteses) da figura 3.4 será apresentada a seguinte tela:

E D I C A O	
OPCAO :	
GRUPOS DE HIPOTEESES	M E N U
	U - Volta tela p/ cima
	D - Volta tela p/ baixo
	G - edita Grupo
	I - Insere grupo
	M - Move grupo
	R - Renomeia grupo
	E - Elimina grupo
	ESC sai
	S T A T U S
	Editando Gpos de Hipoteses

FIGURA 3.5 - "Menu" de Edição de Grupos de Hipóteses. Situação Inicial

Com a opção "I" (Insere) da figura anterior poderemos acrescentar novos grupos de hipóteses, simplesmente digitando os nomes (são considerados no máximo 30 caracteres) desses grupos.

Para o nosso exemplo agruparemos as hipóteses em dois grupos: "Hipóteses de Primeiro Nivel" e "Hipóteses Intermediárias". Geralmente podemos agrupar as hipóteses usando melhores critérios. Por exemplo, SINDROMUS [NICO 87] poderia formar os grupos "SdComAberraçõesCromossômicas", "SdComCraniostenoses", "SdComOsteocondrodisplasias", etc.

A medida que vão sendo criados, os grupos serão apresentados e numerados por ordem de criação (que pode e deve obedecer a algum critério). Desta forma, após a criação do segundo grupo a seguinte tela estará sendo apresentada:

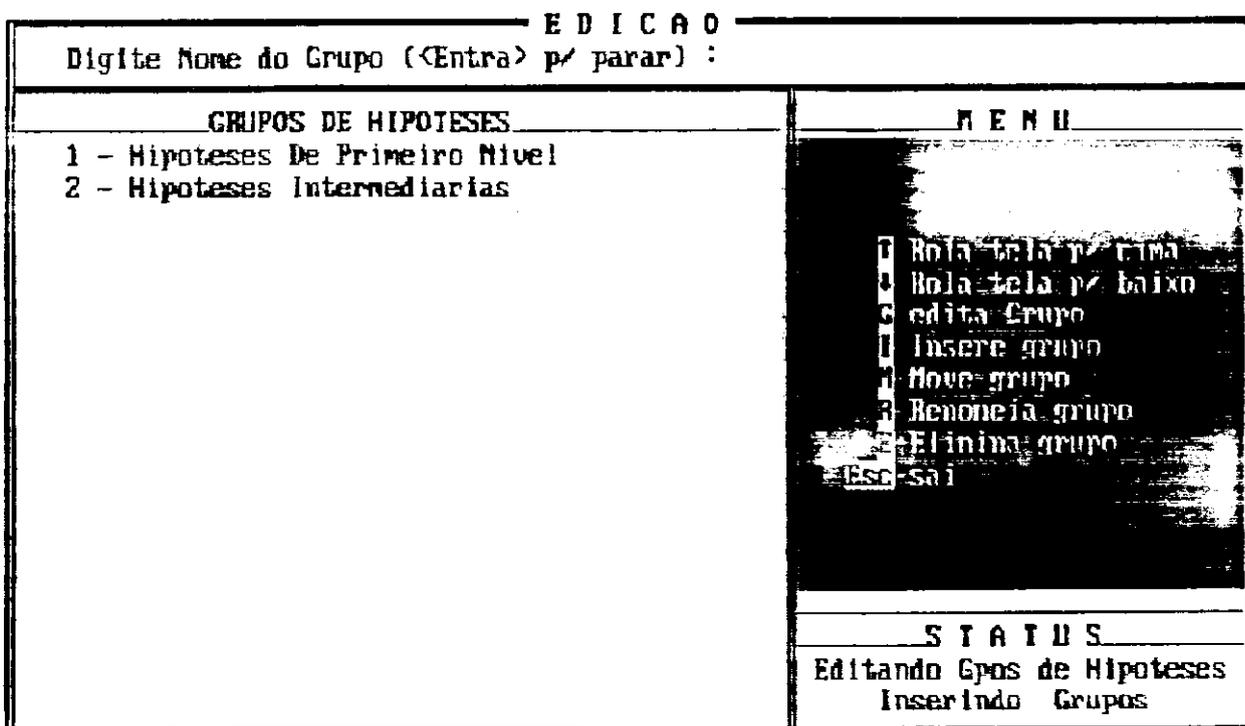


FIGURA 3.6 - "Menu" de Edição de Grupos de Hipóteses. Situação Final

Como não queremos inserir mais nenhum grupo de hipóteses, simplesmente teclamos "Entra" e IntForbayes voltará a pedir nova opção.

Para a inclusão das hipóteses correspondentes a um determinado grupo, bem como para quaisquer outras manipulações com estas hipóteses, deve ser usado a opção "G" (edita Grupo), seguido do número do grupo a ser editado.

Escolhendo, por exemplo, o grupo "Hipóteses de Primeiro Nivel" (número 1), IntForBayes apresentará a seguinte tela:

OPCAO :		E D I C A O	
HIPOTEESES	PROB.PRIORI	M E N U	
		[I] Mola tela p/ cima [J] Mola tela p/ baixo [I] Insere [M] Move [R] Renuncia [E] Elimina [Esc] sai	
		S T A T U S Editando Hipoteses do Gpo. Hipoteses De Primeiro Nivel	

FIGURA 3.7 - "Menu" de Edição de Hipóteses. Situação Inicial

Com a opção "I" (Insere) poderemos informar a IntForBayes quais são as hipóteses que fazem parte do grupo "Hipóteses de Primeiro Nivel". Para cada uma das hipóteses deve ser informado o seu nome e a probabilidade a priori da sua ocorrência no universo dos possíveis consultandos.

Se por exemplo, dentre as pessoas que procuram um especialista em doenças exantemáticas agudas, 1% delas apresentam

Rubéola, dizemos que esta é a probabilidade a priori de ocorrência de Rubéola em um consultando do qual não temos informações.

Supondo que a probabilidade a priori de Sarampo é de 3%, o resultado das inserções do grupo "Hipóteses de Primeiro Nivel" é sintetizado na seguinte tela:

OPCAO :		E D I C A O	
HIPOTESES	PROB. PRIORI	M E N U	
1. Sarampo	3.000000%	Rola tela p/ baixo	
2. Rubéola	1.000000%	Inserc	
		Move	
		Renuncia	
		Elimina	
		Esc Sai	
		STATUS	
		Editando Hipoteses do Gpo.	
		Hipoteses De Primeiro Nivel	

FIGURA 3.8 - "Menu" de Edição de Hipóteses. Situação Final

Concluídas as inclusões do grupo "Hipóteses de Primeiro Nivel", pressionamos a tecla "Esc" (Sai). Com outro "Esc" retornaremos à tela da figura 3.6.

A partir de então, de forma semelhante à utilizada para o grupo "Hipóteses de Primeiro Nivel", introduziremos a sub-hipótese "Infecção das Vias Aéreas Superiores" no grupo "Hipóteses Intermediárias".

As demais opções que aparecem na figura 3.6 são utilizadas para: rolagem das telas de grupos de hipóteses (<SetaCima>, <SetaBaixo>) quando estes ocupam mais de uma tela; Mover um grupo (M) de uma determinada posição para outra qualquer; Renomear um grupo (R); e Eliminar um grupo vazio (E). As opções da tela da figura 3.7 são análogas às da tela da figura 3.6.

Após as inclusões acima, com o uso da tecla "Esc" retornamos à tela da figura 3.4.

A opção "E" (edita grupo de Evidências) desta tela possibilitará, de maneira análoga à descrita para as hipóteses e seus grupos, a edição das evidências e grupos de evidências. Salientamos que para as evidências não serão solicitadas suas probabilidades a priori, mas somente os seus nomes (só serão considerados seus 30 primeiros caracteres).

No exemplo que estamos acompanhando introduziremos o grupo "Nariz e Boca" (com as evidências "Dor Garganta" e "Coriza") e o grupo "Outros" (com a evidência "Febre Alta").

A opção "D" (Define ligações) da figura 3.4 deve ser usada para estabelecer os relacionamentos existentes entre os diversos fatos que constituem a base de conhecimento. A sua escolha implicará na apresentação da seguinte tela:

?

DEFINICAO DE LIGACOES		
<input checked="" type="checkbox"/> Cria Ligacoes	<input type="checkbox"/> Elimina Ligacoes	<input type="checkbox"/> Altera Ligacoes P(F; P)
<input type="checkbox"/> Grava	<input type="checkbox"/> Esc sai	<input type="checkbox"/> seleciona <input type="checkbox"/> muda tela ativa P(F; P)
P A I S	PROB. PRIORI	F I L H O S
Sarampo	3.0000%	Sarampo
Rubeola	1.0000%	Rubeola
Infec Vias Aereas Superiores	9.5000%	Infec Vias Aereas Superiores
Doencas Exantematicas Agudas	50.0000%	Doencas Exantematicas Agudas
		Dor Garganta
		Coniza
		Febre Alta

FIGURA 3.8 - "Menu" Principal de Ligações

A opção "C" (Cria ligações) da figura 3.8 é usada para criar os relacionamentos existentes entre os fatos da base de conhecimento. Sua ativação implicará na apresentação da seguinte tela:

CRIACAO DE LIGACOES		
Cria ligacao entre os fatos indicados		Esc sai P(F; P)
seleciona		→ muda tela ativa P(F; P)
PAIS	PROB. PRIORI	FILHOS
Sarampo	3.0000%	Sarampo
Rubeola	1.0000%	Rubeola
Infec Vias Aereas Superiores	9.5000%	Infec Vias Aereas Superiores
Doencas Exantematicas Agudas	50.0000%	Doencas Exantematicas Agudas
		Dor Garganta
		Coriza
		Febre Alta

FIGURA 3.10 - "Menu" de Criação de Ligações. Situação Inicial

Nas telas das figuras 3.9 e 3.10 um dos seus lados se apresentará de forma destacada (ativa), através da intensidade das letras que aparecem no video e em cores diferenciais se o video do computador for colorido. A este lado chamaremos simplesmente de tela ativa.

Em cada lado da tela existirá um fato que, pelo brilho das letras, se destacará dos demais. Nas figuras deste texto estarão sublinhadas. Qualquer referência a fatos nas telas de definição/criação/eliminação/alteração de ligações será associada aos fatos destacados nestas telas.

Os identificadores que aparecem do lado esquerdo da tela (abaixo de "PAIS") formam o conjunto das hipóteses ou sub-hipóteses (ambas também chamadas fatos dedutíveis, ver 2.2), que

fazem parte da base de conhecimento. A última hipótese mostrada neste lado é um fato especial, criado automaticamente por IntForBayes, cujo nome é exatamente igual ao da base de conhecimento.

Os identificadores do lado direito da tela (abaixo de "FILHOS") formam o conjunto dos fatos que constituem a base de conhecimento. Os identificadores que aparecem acima do nome da base de conhecimento são os mesmos daqueles do lado esquerdo da tela (isto é, os fatos dedutíveis) e os subsequentes são as evidências, também chamadas de fatos perguntáveis (ver 2.2).

Para selecionar um fato, devem ser usadas as teclas "<SetaCima>" ou "<SetaBaixo>", que movimentarão o ativador de fatos ao longo da tela ativa. A seleção de um fato no lado da tela que não se encontra ativa deve ser precedida do uso da tecla "<+>" ou "<->", cuja função é intercambiar os lados ativos e inativos da tela. As teclas "PgDn" e "PgUp" poderão ser utilizadas para avançar ou retroceder uma página, respectivamente, na tela ativa.

Observando a estrutura do conhecimento apresentada na figura 3.3, verificamos que "Febre Alta" é um dos sintomas de "Sarampo", isto é, que "Sarampo" é pai de "Febre Alta". Para informarmos a IntForBayes a dependência entre os fatos acima, a partir da tela da figura 3.10, selecionamos "Sarampo" na lista dos pais e "Febre Alta" na lista dos filhos. Em seguida teclamos "C" (Cria ligação entre os fatos indicados) e IntForBayes solicitará:

(1) a probabilidade de "Febre Alta" (FILHO) dado que

"Sarampo" (PAI) é verdade: $(P(F:P) = 40\%)$.

(2) a probabilidade de "Febre Alta" dado que "Sarampo" é falso: $(P(F:\sim P) = 4\%)$.

Após informarmos estes dados (entrando 40 e 4 nos campos apropriados), IntForBayes apresentará a seguinte tela:

CRIACAO DE LIGACOES		
Qual a probabilidade do FILHO dado PAI :	40	% P(F: P)
Qual a probabil. do FILHO dado nao PAI :	4	% P(F:\~P)
PAIS	PROB. PRIORI	FILHOS
Sarampo	3.00000%	Sarampo
Rubeola	1.00000%	Rubeola
Infec Vias Aereas Superiores	9.50000%	Infec Vias Aereas Superiores
Doencas Exantematicas Agudas	50.00000%	Doencas Exantematicas Agudas
		Dor Garganta
		Coriza
		<u>Febre Alta</u>

FIGURA 3.11 - "Menu" de Criação de Ligações. Situação Final

As outras ligações poderão ser feitas de forma semelhante. Salientamos que as hipótese de primeiro nível devem ser ligadas, obrigatoriamente, à hipótese especial "Doenças Exantemáticas Agudas". As probabilidades condicionais usadas nestas ligações não serão utilizadas por IntForBayes, portanto, caso não sejam conhecidas, poderão ser utilizados valores arbitrários.

Após a criação das ligações desejadas, a tecla "Esc" (sai), da tela da figura 3.10, deve ser utilizada para voltar à

tela da figura 3.9, atualizada.

As opções "E" (Elimina ligações) e "A" (Altera ligações) devem ser usadas, respectivamente, para 1) eliminar uma ligação anteriormente criada ou 2) fazer alguma alteração numa ligação ou nos fatos que ela envolve. Além disso, o uso de uma dessas opções é indicado caso se queira simplesmente visualizar quais são os filhos de um pai especificado (ativado).

Em qualquer uma das telas e subtelas de definição de ligação (figuras 3.9 a 3.11), caso exista uma ligação entre os fatos destacados, as probabilidades condicionais desta ligação aparecerão ao lado de $P(F:P)$ e de $P(F:\sim P)$.

A opção "Esc" (sai), da tela da figura 3.9, deve ser usada para retornar à tela da figura 3.4.

A opção "A" (Altera ponto de satisfação), desta tela, deve ser utilizada para aumentar (diminuir) o número de perguntas e a confiabilidade das respostas, através da diminuição (aumento) do ponto de satisfação a partir do qual, se a importância de um fato ficar abaixo desse ponto, IntForBayes dará por terminada a busca de informações que influenciam esse fato (em outras palavras, esse ponto de satisfação é um "critério de poda"). Ao ser ativada a opção "A", será apresentada a seguinte tela:

Ponto de Satisfacao Atual = 0.050000 Novo Ponto de Satisfacao =	
Tecla apenas <Enter> para nao alterar	
1.0	nao faz perguntas, minima precisao dos resultados
0.0	faz todas as perguntas, maxima precisao dos resultados

FIGURA 3.12 - Tela de Alteração do Ponto de Satisfação

Se não escolhermos um valor para o ponto de satisfação na edição inicial criativa de um S.E., então IntForBayes atribuir-lhe-á um valor "default" de 0.05000 (modificável na instalação de IntForBayes. Ver apêndice A e também a seção 6.2). Mas provavelmente será melhor que, naquela edição, fixemos o ponto de satisfação num valor que os subjetivismo e experiência, do engenheiro de conhecimento e dos especialistas, apontem como o mais apropriado para aquele domínio de conhecimento (no caso de dúvidas, é melhor errar por fixar o ponto de satisfação mais próximo de 0 do que errar por fixá-lo mais longe).

Durante cada consulta o ponto de satisfação pode ser reajustado à vontade, devendo ser aumentado se perguntas demais estiverem sendo feitas, e diminuído se precisão de menos estiver sendo obtida; mas, na próxima reinicialização do S.E., o ponto de satisfação voltará ao valor fixado na última edição. Assim, aconselhamos que: 1) à medida que formos adquirindo maior experiência com o uso do S.E., usemos o editor para readequar (de

forma duradoura) o ponto de satisfação; 2) quando, ao obtermos os resultados de uma consulta, não os julgarmos aceitavelmente precisos, diminuamos (provisoriamente) o ponto de satisfação; 3) quando temermos que o S.E. fará perguntas demais num caso que nos parece simples e sem necessitar muita precisão, aumentemos (provisoriamente) o ponto de satisfação.

Para manter o ponto de satisfação mostrado na tela da figura 3.12, deve ser usada a opção "Entra", retornando assim à tela da figura 3.4.

Notemos que o ponto de satisfação pode ser definido em três ocasiões diferentes: na configuração de IntForBayes, na edição de um S.E. e numa consulta.

Uma melhor explicação do cálculo das importâncias e do papel do ponto de satisfação ultrapassa o escopo deste tutorial e deve ser vista em 2.3, 5.1, 5.2 e 5.3.

Concluídas as edições da base de conhecimento "Doenças Exantemáticas Agudas", deve ser usada a opção "G" (Grava arquivos) para registrar em disco as modificações que já estão processadas na memória interna. Aconselhamos gravar a base de conhecimento a intervalos regulares (digamos, de uns 15 minutos), para não perder muito se houver algum problema como, por exemplo, interrupção no fornecimento de energia elétrica.

A opção "Esc" (sai) da tela da figura 3.4 deve ser usada tanto para retornar à tela da figura 3.2, como para daí retornar à tela da figura 3.1. Neste ponto o sistema estará apto a

utilizar a base de conhecimento, recém criada, para realizar consultas, como será visto a seguir.

3.2.2.2 Executando um pequeno S.E.

Para processar consultas com o conhecimento de alguma base de conhecimento editada por IntForBayes, o usuário deve utilizar a opção "X" (eXecuta base de conhecimento) da tela da figura 3.1, fazendo com que IntForBayes apresente a seguinte tela:

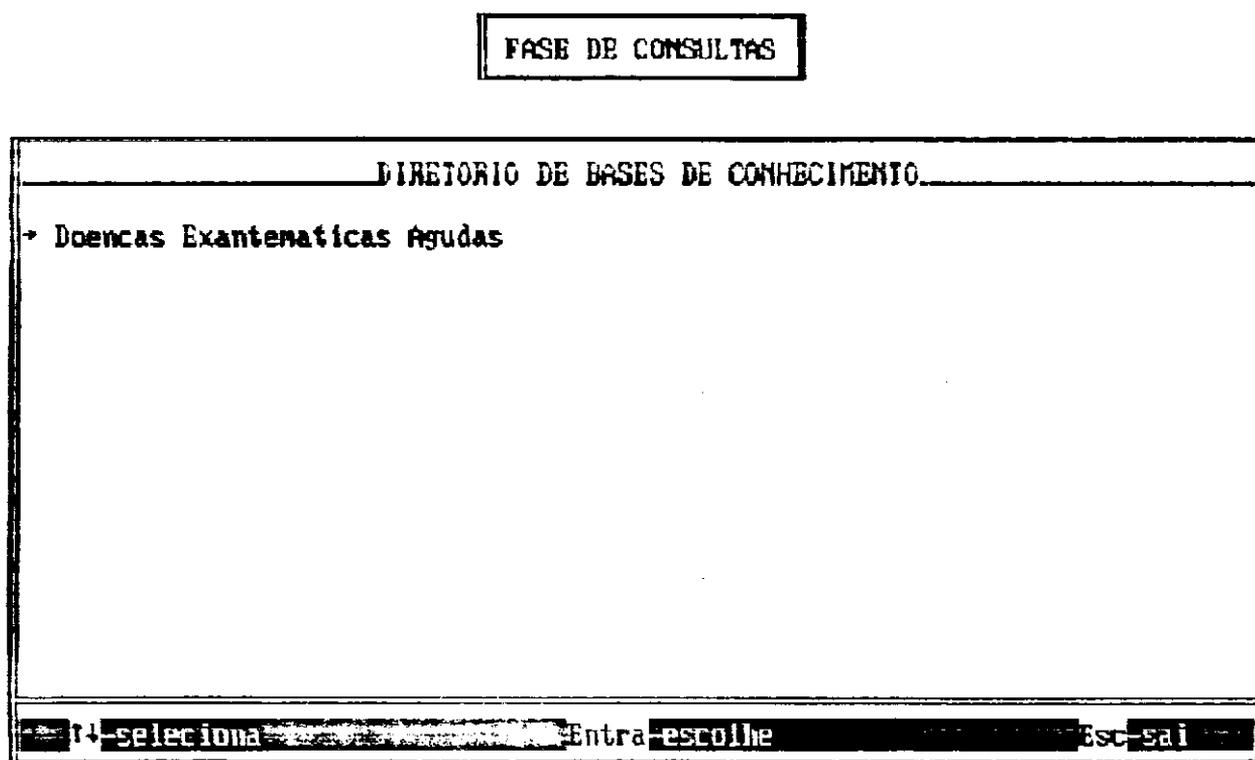


FIGURA 3.13 - "Menu" de Bases Existentes

A tecla "Entr" (Escolhe) desta tela deve ser pressionada para escolher a base de conhecimento indicada (no momento só existe a base do exemplo descrito na seção anterior). Feito isto, IntForBayes solicitará o nome do consultando (só serão considerados seus 30 primeiros caracteres). Informado o nome do

consultando (Lampião Filho), IntForBayes apresentará a seguinte tela:

Nome do consultando: Lampiao Filho

Consultando nao cadastrado.
Deseja cadastrá-lo (s/n)?

FIGURA 3.14 - Tela de Identificação/cadastramento de Consultandos

Para IntForBayes o consultando acima não faz parte do seu cadastro de consultandos. Para prosseguir com esta consulta o usuário deve cadastrá-lo, respondendo com "s" (sim) à pergunta da tela acima. Após o cadastramento de um consultando IntForBayes apresentará a seguinte tela:

Edite comentarios (sobre o consultando), sem efeito nas inferencias:

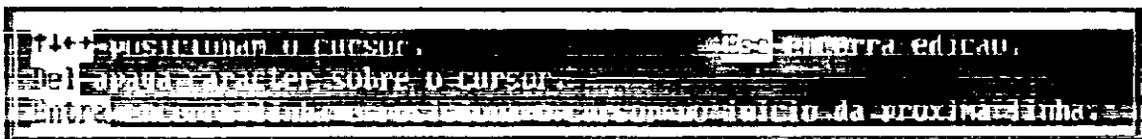
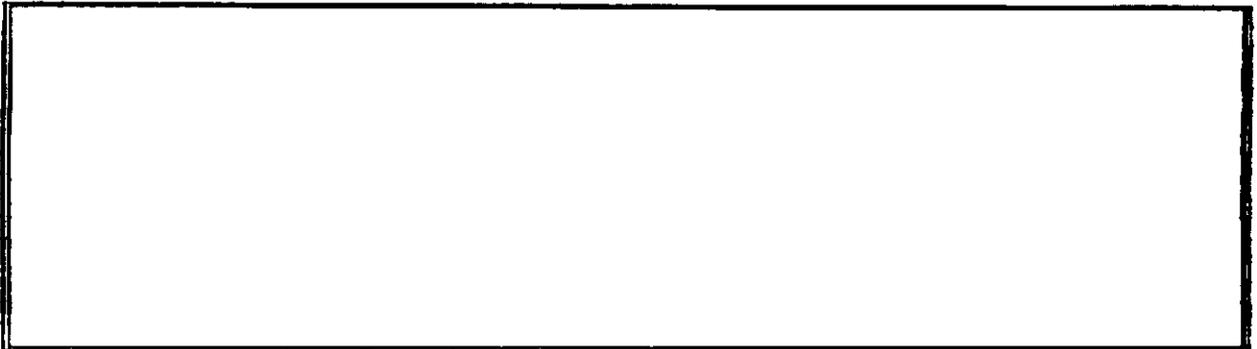


FIGURA 3.15 - Tela de Edição de Comentários

Nesta ocasião o usuário pode utilizar o espaço disponível na tela para colocar os comentários sobre o consultando (isto é, informações que, mesmo sem influenciarem as inferências, sejam julgadas convenientes, tais como, nome, endereço, local e data de nascimento, classificação como pagador, classificação como seguidor das recomendações, etc). Se o consultando já tivesse sido cadastrado em uma consulta anterior, os comentários editados na última consulta estariam sendo apresentadas na tela acima e poderia ser feita qualquer atualização ou adição de novos comentários.

A tecla "Esc" (sai), da tela da figura 3.15, deve ser usada após a conclusão da edição dos comentários sobre o consultando. Com sua ativação IntForBayes apresentará a tela:

Sistema Especialista em Doenças Exantemáticas Agudas

HIPÓTESES	GRAU DE CERTEZA		
	ATUAL	MÍNIMO	MÁXIMO
+ Sarampo	0.000	-3.010	4.256
Rubeola	0.000	-4.908	4.771

1 - seleciona	2 - entra	3 - investiga hipótese	4 - mostra conteúdo de resultado
5 - oferece/corrigir dados	6 - altera ponto de satisfação	7 - esc. sair	

FIGURA 3.16 - Situação Global das Hipóteses. Estado Inicial

Se o consultando fosse antigo e se tivesse anteriormente gravado alguma consulta, antes da apresentação desta tela, IntForBayes perguntaria se deveria continuar com a consulta anterior ou se iniciaria uma nova consulta. Caso fosse solicitada a continuação da consulta anterior, seriam transportados para a memória todos os fatos informados na última consulta, deduzidos todos os fatos possíveis e restaurado integralmente o quadro global apresentado anteriormente; caso contrário, IntForBayes esqueceria a consulta anterior, considerando a consulta corrente como sendo uma nova consulta.

As hipóteses que aparecem na tela da figura 3.16 são as hipóteses de primeiro nível, ou seja, são aquelas que só têm por pai "Doenças Exantemáticas Agudas" e que, assim, podem ser escolhidas para investigação ou para explicação dos resultados

obtidos. Para cada hipótese é apresentado seu grau de certeza atual e os valores extremos (mínimo e máximo) que ele pode atingir. O grau de certeza de uma hipótese mede a crença nela, com base nos dados fornecidos. Sua faixa de valores varia de -5 (descrença absoluta na hipótese) a +5 (crença absoluta na hipótese) e o 0 corresponde à indecisão absoluta. As hipóteses são apresentadas, inicialmente, em ordem decrescente de suas probabilidades a priori e, após o fornecimento de informações, em ordem decrescente dos graus de certeza atuais.

A opção "0" (Oferece/corrige dados) da tela da figura 3.16 deve ser usada para:

- 1) informar os fatos pertinentes ao consultando e que sejam consequentes de relatos espontâneos ou obtidos por observações que saltem à vista, ou sejam obtíveis, com custo (em tempo, dinheiro, incômodo, etc) muito baixo; ou
- 2) corrigir informações fornecidas ao longo da consulta.

No nosso exemplo suponhamos que o consultando se queixe espontaneamente apenas de dor de garganta. Para informar isto a IntForBayes deve ser usada a opção "0" (Oferece/corrige dados) da tela da figura 3.16, fazendo com que seja apresentada a seguinte tela:

Fase de Oferta Voluntária ou Correção de Informações

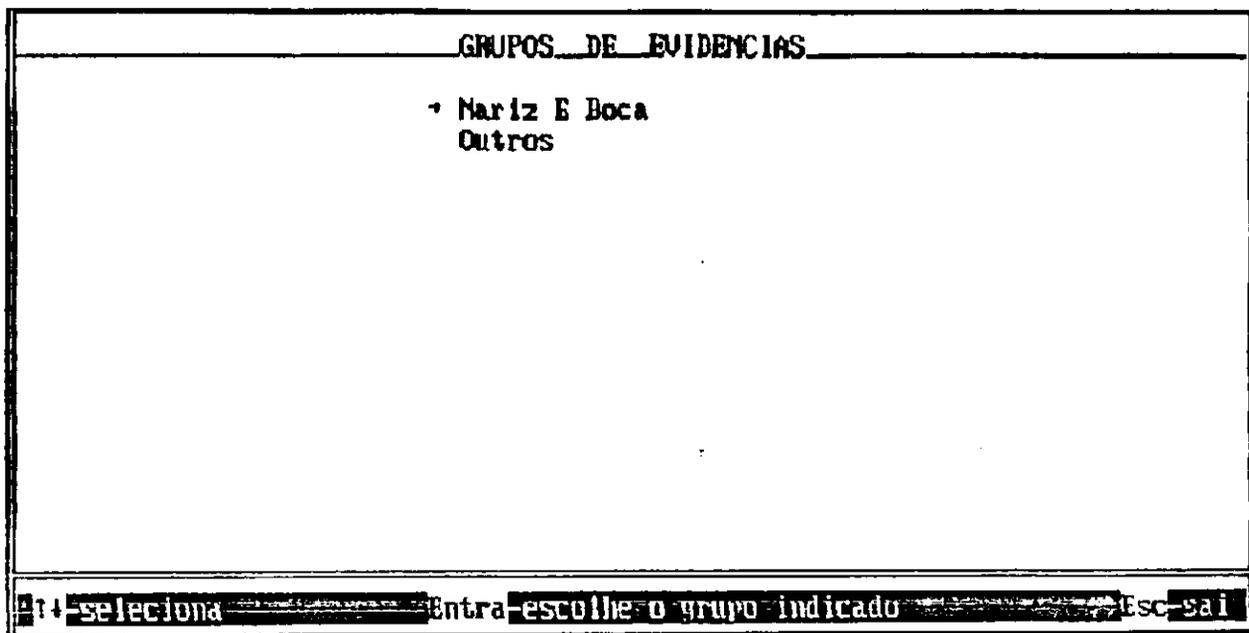


FIGURA 3.17 - "Menu" de Grupos Voluntariáveis

Com a escolha do grupo "Nariz e Boca" IntForBayes apresentará a seguinte tela:

Fase de Oferta Voluntária ou Correção de Informações

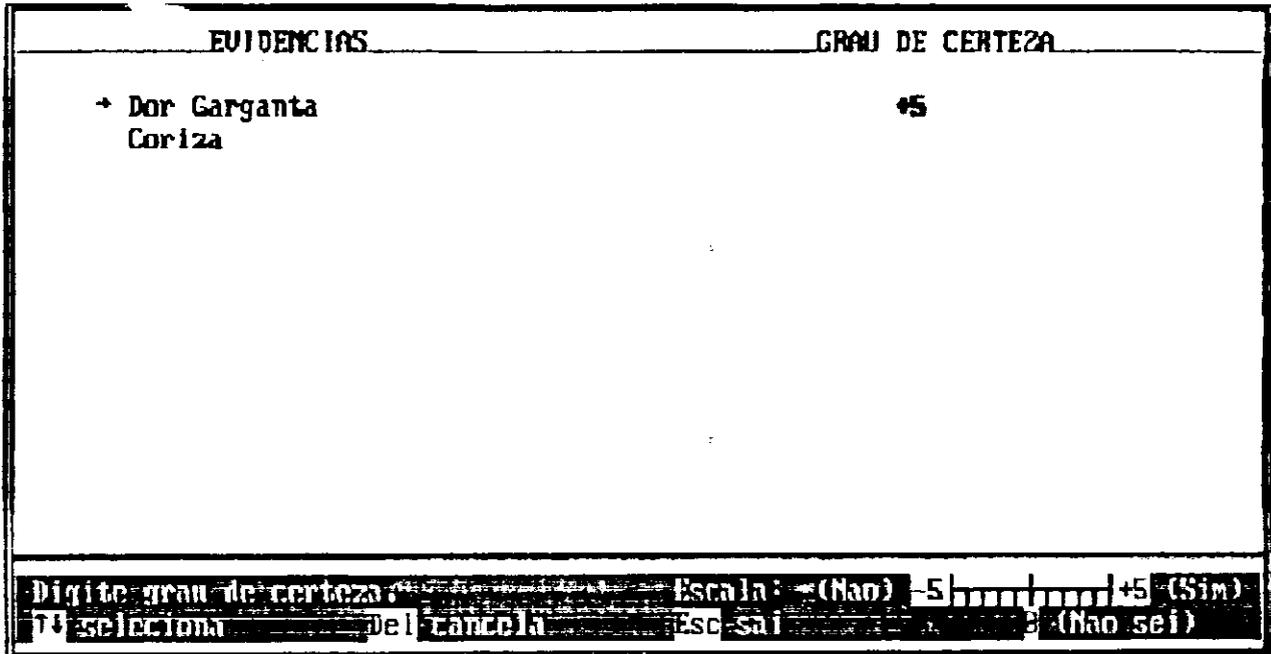


FIGURA 3.18 - "Menu" de Evidências Voluntariáveis

Dentre as evidências do grupo escolhido, selecionamos "Dor de Garganta" e fornecemos o grau de certeza +5 (certeza absoluta), conforme já está mostrado na tela da figura acima. A cada grau de certeza atribuído a uma evidência, IntForBayes imediatamente propaga esta informação através da base de conhecimento, atualizando todos os graus de certeza das hipóteses direta ou indiretamente relacionadas com a evidência. Ao invés do processamento ser feito só ao final da coleta de informações, deixando o usuário à espera por longos segundos, ele é diluído e feito um pouquinho a cada informação recebida, tornando o sistema muito mais agradável ao usuário.

A opção "Del" (cancela) da tela da figura 3.18 deve ser usada para retornar o status do grau de certeza de uma evidência

à condição de "não informado" (status original). Notemos que "não informado" e "informado que não sei" são diferentes. No caso de alteração do valor de um grau de certeza, basta simplesmente digitar o novo valor por cima do valor a ser alterado. Neste caso, IntForBayes anulará todos os efeitos diretos e indiretos do grau de certeza anteriormente informado e propagará a nova informação ao longo de toda a base de conhecimento.

Após o oferecimento das informações do grupo "Nariz e Boca", a tecla "Esc" deve ser usada para retornar à tela da figura 3.17. Nessa tela a tecla "Esc" deve ser novamente usada para sair da fase de oferta/correção de informações. Feito isto, IntForBayes apresentará a nova situação das hipóteses de primeiro nível refletindo as informações fornecidas nessa fase:

Sistema Especialista em Doenças Exantemáticas Agudas

HIPÓTESES	GRAU DE CERTeza		
	ATUAL	MÍNIMO	MÁXIMO
+ Sarango	1.572	0.012	4.256
Rubeola	1.236	-3.740	4.771

↑ seleciona ← Entra Investiga hipótese → Mostra Como obteve resultado
 ↓ Oferece/corre dados ← Altera linear de importância → Esc sai

FIGURA 3.19 - Situação Global das Hipóteses. Estado 2

As hipóteses são ordenadas por ordem decrescente dos graus de certeza atuais e a primeira hipótese fica ativada como pronta para ser escolhida para investigação. No entanto, supondo que no momento "Rubéola" é de determinação mais urgente (ou preferida por qualquer outro motivo), mandamos IntForBayes investigar esta hipótese, e a seguinte tela será apresentada:

Fase de Perguntas

HIPOTESE EM ANALISE	GRAU DE CERTEZA		
	ATUAL	MINIMO	MAXIMO
Rubeola	1.236	-3.748	4.771
<p>COM QUE GRAU DE CERTEZA VOCE ACHEDITA EM "Febre Alta"?</p>			
<p> <input type="text"/> Digite grau de certeza. Escala: (Nao) -5 (Sim) <input type="text"/> responde Para que a pergunta Esc: sai 3 (Nao sei) </p>			

FIGURA 3.20 - Situação Global das Hipóteses. Estado 3

Nesta fase (fase de perguntas) IntForBayes irá tentar provar a veracidade ou falsidade da hipótese investigada, com a obtenção de novas informações resultantes de perguntas feitas ao usuário. Estas perguntas serão escolhidas de modo que, a cada momento, seja feita a "melhor" (a "mais inteligente") pergunta para atingir o objetivo. A medida que as perguntas vão sendo feitas, IntForBayes manterá constantemente atualizados na tela os

graus de certeza (atual, mínimo e máximo) da hipótese investigada, de modo que o usuário pode sentir sua tendência, bem como a contribuição de cada pergunta para o resultado apresentado.

A sequência de perguntas se dará até que 1) a importância da hipótese investigada se torne menor que o ponto de satisfação estabelecido; ou 2) não existam mais perguntas com respostas capazes de influenciar a hipótese investigada; ou 3) o usuário já esteja satisfeito com a precisão dos resultados apresentados e tecla "Esc".

Na tela da figura 3.20 IntForBayes informa que está investigando "Rubéola", que possui atualmente um grau de certeza de 1,236 e que, mesmo que todas as evidências que ainda não foram informadas e que interferem no resultado de "Rubéola" sejam respondidas com os valores de graus de certeza que mais contribuam para a sua negação (confirmação), o grau de certeza alcançado por essa hipótese será de -3,740 (4,771). Além disso, a evidência mais inteligentemente escolhível no momento, para IntForBayes alcançar com menos perguntas sua meta (provar ou desprovar "Rubéola"), é "Febre Alta", cujo grau de certeza o usuário deve digitar nesta ocasião.

Além de um grau de certeza válido na faixa de -5 (descrença absoluta) a +5 (crença absoluta), o usuário pode também: 1) utilizar a tecla "P" ("Para que a pergunta?") * para pedir a IntForBayes sua justificativa para a formulação da pergunta

* Ainda não implementada.

atual; ou ainda 2) a tecla "Esc" (sai), para interromper a sequência de perguntas, fazendo com que sejam apresentados os resultados parciais das hipóteses de primeiro nível, obtidos com as informações fornecidas até o momento.

Supondo que o consultando do nosso exemplo nos pareça, com razoável certeza, ser caracterizável pela evidência febre alta, respondemos à pergunta feita na tela da figura 3.20 informando que o grau de certeza de "Febre Alta" é +3. Com isto, IntForBayes apresentará a seguinte tela:

Sistema Especialista em Doenças Exantemáticas Agudas

H I P O T E S E S	G R A U D E C E R T E Z A		
	ATUAL	MINIMO	MAXIMO
→ Rubéola	4.573	4.429	4.635
Sarampo	3.779	3.468	3.925

seleciona Entra-investiga hipótese mostra como obter resultado
 Oferece/corrigir dados Altera nível de importância Esc-sai

FIGURA 3.21 - Situação Global das Hipóteses. Estado 4

Como estamos usando um ponto de satisfação relativamente baixo e como IntForBayes procura fazer as perguntas sempre na ordem decrescente das suas influências para o estabelecimento dos resultados, as evidências que não são consideradas na

investigação de uma hipótese pouco influenciariam no resultado final. No nosso exemplo: 1) Em números aproximados, o grau de certeza de "Rubéola" poderia passar dos atuais 4,6 para no mínimo 4,4 e no máximo 4,7. 2) Através da figura 3.3 percebemos que a evidência não considerada no exemplo foi "Coriza". 3) Supondo que não estamos satisfeitos com a precisão dos resultados apresentados, podemos, com o uso da tecla "A" (Altera ponto de satisfação), reduzir o ponto de satisfação dos atuais 0.05 para 0.0001, por exemplo. 4) Feito isto e selecionando novamente "Rubéola" para uma investigação, IntForBayes apresentará a seguinte tela:

Fase de Perguntas

HIPOTESE EM ANALISE	GRAU DE CERTEZA		
	ATUAL	MINIMO	MAXIMO
Rubeola	4.573	4.429	4.635
COM QUE GRAU DE CERTEZA VOCE ACHEDITA EM "Coriza"?			
<div style="display: flex; justify-content: space-between;"> Digite grau de certeza: Escala: (Não) -5 (Sim) </div> <div style="display: flex; justify-content: space-between;"> responde para que a pergunta Esc. sai: (Não sei) </div>			

FIGURA 3.22 - Tela de Perguntas

5) Supondo que o nosso consultando definitivamente não apresente "Coriza", respondemos à pergunta acima com grau de certeza -5 (não) e IntForBayes apresentará a seguinte tela:

Sistema Especialista em Doenças Exantemáticas Agudas

HIPÓTESES	GRAU DE CERTEZA		
	ATUAL	MINIMO	MAXIMO
+ Rubéola	4.429	4.429	4.429
Sarampo	3.468	3.468	3.468

T: seleciona; E: entra; I: investiga; hipotese; C: mostra; Como; obtive; resultado;
 O: Oferece/corrigir dados; A: altera; Liniar; de; importancia; Esc: sai

FIGURA 3.23 - Situação Global das Hipóteses. Estado 5

6) Pela igualdade dos graus de certeza (atual, mínimo e máximo), constatamos que todas as perguntas relacionadas com as hipóteses acima foram respondidas. 7) Dentro das limitações da base de conhecimento, interpretamos o resultado como diagnosticando (ou melhor, classificando) "Rubéola" como a hipótese mais plausível, embora "Sarampo" também possa ser aceitável.

Arrependendo-nos, por exemplo, da informação que prestamos sobre "Dor de Garganta", entramos na fase de oferta/correção de dados (opção "O") para informar que o grau de certeza desta evidência é -5 (não). Ao sairmos dessa fase, IntForBayes apresentará a seguinte tela:

Sistema Especialista em Doenças Exantemáticas Agudas

HIPOTESSES	GRAU DE CERTEZA		
	ATUAL	MINIMO	MAXIMO
→ Sarampo	0.413	0.413	0.413
Rubéola	0.296	0.296	0.296

T - seleciona E - Entra I - investiga hipótese C - mostra Como obteve resultado
 O - Oferece/corrigir dados A - Altera ponto de satisfação Esc - sai

FIGURA 3.24 - Situação Global das Hipóteses. Estado 6

A opção "C" ("Como obteve o resultado?") * da tela da figura acima deve ser usada sempre que o usuário deseje saber a justificativa (explicação) da linha de raciocínio utilizada para atingir os resultados apresentados.

Para concluir a consulta atual, deve ser utilizada a opção "Esc" (sai), da tela da figura 3.24, fazendo com que seja apresentada a seguinte tela:

* Ainda não implementada.

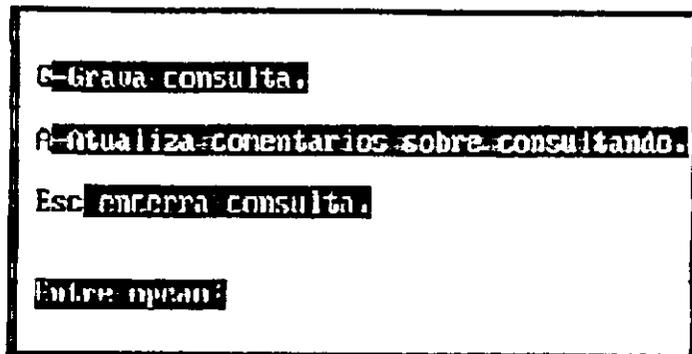


FIGURA 3.25 - "Menu" de Finalização de Consultas

A opção "G" (Grava consulta) deve ser utilizada para guardar as informações obtidas durante a consulta corrente. Desta forma, se desejadas, todas as informações (comentários + graus de certeza das evidências) fornecidas e os resultados obtidos poderão ser recuperados na próxima consulta do consultando correspondente.

A opção "A" (Atualiza comentários sobre o consultando) deve ser usada para retornar à tela da figura 3.15 e complementar ou alterar comentários e anotações sobre o consultando.

A opção "Esc" (sai) deve ser usada para finalizar definitivamente a consulta atual, fazendo com que IntForBayes retorne à tela da figura 3.14 pedindo o nome do próximo consultando. Após as consultas desejadas, a opção "Esc", desta tela, deve ser usada para finalizar a sessão com a presente base de conhecimento, voltando assim à tela da figura 3.13. Se desejado, podemos iniciar uma sessão com outra (ou a mesma) base de conhecimento, sendo deve ser usada a tecla "Esc", da tela da

figura anterior, fazendo com que seja apresentada a tela da figura 3.1, na saída do módulo de execução. Para sair de IntForBayes, retornando ao sistema operacional, deve ser usada a tecla "Esc" desta tela.

3.3 ALGUNS CONSELHOS ÚTEIS

Na elaboração de bases de conhecimento é conveniente e desejável que obedeçamos a alguns critérios:

- 1) Evitemos a utilização de fatos muito correlacionados (por exemplo, Febre e Temperatura Alta) ou que sejam muito complementares (por exemplo, Sexo Masculino e Sexo Feminino: a negação de um deles torna automática a afirmação do outro, portanto a utilização de apenas um desses fatos é suficiente).
- 2) Durante a edição, somente se não dispusermos de estatísticas devemos usar valores estimados ou subjetivos para as probabilidades. Neste caso usemos de muita ponderação.
- 3) No processo de desenvolvimento de um S.E. e, principalmente, no seu uso, adotemos o seguinte procedimento geral, que visa aumentar as produtividades nossa e a da máquina:

```

defina a base de conhecimento e o ponto de satisfação "default"
para todas as sessões com a base;
enquanto desejado {rode uma sessão com a base de conhecimento}:
| enquanto desejado {rode uma consulta}:
| | enquanto desejado {rode um ciclo da consulta}:
| | | voluntarie as evidências que lhe saltem à vista, lhes foram
| | | doadas, você obtem com baixissimo custo, ou sua intuição
| | | lhe pressiona a oferecer
| | | se satisfeito com os resultados então abandone a consulta
| | | escolha uma hipótese de primeiro nível para investigar
| | | enquanto o sistema lhe fizer pergunta e a diferença
| | |  $G_{CMax} - G_{CMin}$  (da hipótese escolhida) não lhe satisfaz:
| | | então responda à pergunta
| | | fim
| | | se satisfeito com os resultados então abandone a consulta
| | | se existe outra hipótese de primeiro nível mais promissora
| | | porém muito indefinida (i.é., com  $G_{CMax} - G_{CMin}$ 
| | | insatisfatório)
| | | então troque a hipótese a investigar
| | | senão (i.é., a hipótese escolhida ainda é a sua
| | | preferida, mas IntForBayes não conseguiu determiná-la
| | | bem): diminua o ponto de satisfação, só para esta
| | | sessão
| | fim
| fim
fim
se desejado, altere a base de conhecimento e o ponto de
satisfação "default" para todas as sessões com a base de
conhecimento

```

4. PROJETO E IMPLEMENTAÇÃO

A linguagem de programação utilizada na implementação de IntForBayes foi Pascal. A opção por esta linguagem justifica-se pela sua disponibilidade para quase todas as máquinas de interesse, sua disponibilidade local, sua elegância e popularidade (especialmente nas universidades), a boa estrutura e legibilidade de seus programas e seu atendimento às necessidades de eficiência e estruturas de dados desejadas. Foi usada a versão 5.0 da Borland mas, favorecendo a transportabilidade do sistema, restringimo-nos o quanto pudemos ao Pascal padrão definido por [JENS 74].

A seguir, apresentaremos as principais estruturas de dados e os principais algoritmos de IntForBayes.

4.1 Estruturas de Dados

Eficiência na execução de um S.E. é geralmente mais importante que na edição. Na execução a base de conhecimento é imutável e só precisamos de estruturas de dados estáticas. Na edição, poderíamos usar gerenciamento estático ou gerenciamento dinâmico de memória, mas este último implicaria num código mais complexo e lento, prejudicando a execução. Assim, optamos por trabalhar com estruturas de dados estáticas, porém com limites facilmente alteráveis de acordo com as necessidades, utilizando as constantes definidas no apêndice A.

As principais estruturas de dados de IntForBayes são

constituídas das seguintes tabelas:

Bases (1..LinBases)

(pesquisa sequencial)

<u>Nome</u>	Codigo	PtoSatisf
-------------	--------	-----------

Rede (1..LinRede)

Rede + LstPais + LstFilhos: alocação sequencial, juntas implementam um grafo, pesquisável como árvore)

Import	ApLstPais	Grau de Certeza	TagCar	Nome do fato perg					
				Nome do Fato Ded	Teto Alcancavel	Piso Alcancavel	Chance a Priori	Chance Atual	ApLstFilhos

LstPais (1..LinLstPais)

ApPal	Fator do Sim	Fator do Nao	ApProxPal	FAuxChance	FAuxTeto	FAuxPiso
-------	--------------	--------------	-----------	------------	----------	----------

LstFilhos (1..LinLstFilhos)

ApFilho	ApProxFilho
---------	-------------

Grupos de Hipoteses (1..LinLstGpos)

(pesquisa sequencial)

<u>Nome</u>	Ap InicGpo
-------------	------------

LstHips (1..LinLstHips)

(alocação sequencial)

ApHipotese

Grupos de Evidencias (1..LinLstGpos)

(pesquisa sequencial)

<u>Nome</u>	Ap InicGpo
-------------	------------

LstEvids (1..LinLstEvids)

(alocação sequencial)

ApEvidencia

FIGURA 4.1 - Estruturas de Dados de IntForBayes

Sublinhamos os nomes dos campos normalmente usados como chave principal nas pesquisas.

A tabela Bases identifica as bases de conhecimento que estão disponíveis para consultas ou manutenções. A identificação interna de cada base é feita pelo seu código e para o usuário pelo seu nome. Nesta tabela também é guardado o Ponto de Satisfação definido na edição. A alteração do Ponto de Satisfação, permitida na fase de consultas, não afetará o seu valor nesta tabela.

A tabela Rede contém as informações que caracterizam os fatos constituintes da base de conhecimento. Cada nodo da rede de conhecimento é representado por uma entrada desta tabela. O primeiro campo representa a importância do fato, calculado de acordo com a fórmula (11) de 2.3. ApLstPais é um índice apontando para o início da lista dos pais do fato, na tabela LstPais. O valor de grau de certeza varia de -1.0 (descrença absoluta no fato) a +1.0 (crença absoluta no fato). Ele é inicializado com 0.0 se o fato for dedutível (hipótese), significando ignorância absoluta do fato à luz dos dados fornecidos; e com +6.0 (indicador de não informado) se o fato for perguntável (evidência), significando ignorância absoluta por ainda não ter sido voluntariado/corrigido nem perguntado-respondido. A tabela Rede é constituída por dois tipos de registros, um referente aos fatos perguntáveis (TagCar igual a "P") e o outro referente aos fatos dedutíveis (TagCar igual a "D"). Nos fatos perguntáveis, além dos campos anteriores há o do seu nome. Nos fatos dedutíveis, além dos campos anteriores há o do seu nome, os dos

Teto Alcançável e Piso Alcançável (que contém os valores máximo e mínimo que a chance atual poderá atingir), o da Chance a Priori (que é a chance que corresponde à probabilidade a priori do fato), o da Chance Atual (que é inicializado com o valor da chance a priori e é atualizado de acordo com as fórmulas apresentadas em 2.2) e o campo ApLstFilhos que é um índice apontando para o início da lista de filhos do fato na tabela LstFilhos. A raiz da árvore (com o campo Nome igual ao nome da base de conhecimento) fica na primeira posição de Rede.

Na tabela LstPais, cada registro corresponde a uma ligação filho-pai entre dois fatos da rede de conhecimento. ApPai é um inteiro cujo valor corresponde ao índice do fato "Pai" (envolvido na ligação) na tabela Rede. Fator do Sim e Fator do Não são calculados em função das probabilidades condicionais entre os fatos envolvidos na ligação (ver 2.2). ApProxPai é um inteiro cujo valor corresponde ao índice da tabela LstPais da entrada da ligação para o próximo pai do corrente filho. FAuxChance, FAuxTeto e FAuxPiso guardam os valores anteriores dos fatores multiplicativos usados, respectivamente, no cálculo da chance, teto alcançável e piso alcançável do fato "Pai" envolvido na ligação. Os valores iniciais e as fórmulas usadas na atualização destes fatores são vistos em detalhes em 2.2.

Ao ser criada uma ligação de dependência entre dois fatos da base de conhecimento, são criados um registro na tabela LstPais e um registro na tabela LstFilhos. Portanto, para cada registro da tabela LstPais existe um equivalente na tabela

LstFilhos. O campo ApFilho da tabela LstFilhos é um inteiro cujo valor corresponde ao índice do fato "Filho" (envolvido na ligação) na tabela Rede. ApProxFilho é um inteiro cujo valor corresponde ao índice na tabela LstFilhos da entrada da ligação para o próximo filho do presente pai. As tabelas Rede, LstPais, LstFilhos têm alocação e pesquisa sequenciais. Mas, juntas, implementam um grafo direcionado e acíclico, pesquisável como árvore nas duas direções (raiz → folhas, folha → raiz).

Em IntForBayes os fatos (hipóteses ou evidências) podem ser agrupados, usando critérios que o construtor da base julgue adequados, para uma maior facilidade e rapidez no uso da interface com a máquina, principalmente no casos de bases de conhecimento com elevado número de fatos. As tabelas Grupos de Hipóteses e Grupos de Evidências são utilizadas para definir tais agrupamentos. Os campos Nome e ApInicGpo da tabela Grupos de Hipóteses (Grupos de Evidências) são, respectivamente, o nome do grupo e um inteiro cujo valor corresponde ao índice na tabela LstHips (LstEvids) onde se encontra o inteiro correspondente ao índice para a primeira hipótese (evidência) do grupo. O índice para a última hipótese (evidência) de um grupo é definido pelo decremento de uma unidade no valor da ApHipótese (ApEvidência) da primeira hipótese (evidência) do grupo seguinte. O último grupo da tabela Grupo de Hipóteses (Grupo de Evidências) é criado automaticamente apenas para servir de identificação (flag) do último fato do penúltimo grupo. Salientamos que os agrupamentos têm efeito apenas na edição e no voluntariamento/correção de dados, mas não na rede (constituída das tabelas Rede, LstPais e

LstFilhos), nem no processo de inferências.

A tabela LstHips (LstEvids) contém, em registros contíguos, os índices para as hipóteses (evidências) dos grupos correspondentes. O campo ApHipótese (ApEvidência) guarda estes índices.

Sob o diretório raiz de IntForBayes existe o arquivo ARQ_CONS. Sob cada diretório correspondendo a uma base específica de conhecimento existe, para cada consultando, 0 ou 1 arquivos cujo prefixo é o código do consultando e cuja extensão é ".PRG". ver figura 4.2.

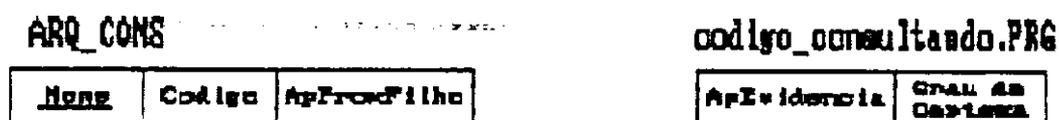


FIGURA 4.2 - Arquivos Complementares

O arquivo ARQ_CONS contém um registro para cada consultando e cada registro contém o nome do consultando, seu código e o campo ApProxCons que é um índice que auxilia na busca de um determinado consultando. Inicialmente é fixado um número provisório de registros (TamTabCons, definido nas constantes de configuração de IntForBayes - apêndice A), que a tabela conterá. Para pesquisar um consultando nesta tabela, objetivando cadastrá-lo ou confirmar seu cadastramento, é usada uma função de hash que utiliza os caracteres que compõem o seu nome e retorna o índice na tabela do possível registro deste consultando. No caso de conflito (coincidência) no acesso, o campo ApProxCons, do tipo inteiro, servirá para indicar o próximo registro a ser

pesquisado, numa lista além do alcance da função de hash. Este processo é utilizado para garantir uma maior velocidade de acesso na pesquisa de um consultando.

No arquivo `codigo_consultando.PRG` é sintetizada a última consulta do corrente consultando. `ApEvidência` e `Grau de Certeza` são, respectivamente, o índice (na tabela `Rede`) e o grau de certeza da evidência referenciada pelo consultando na consulta correspondente. Ao invés de guardar todas as informações contidas nos fatos informados e nos deduzidos (o que implicaria na utilização de muita memória) para reativação posterior de uma consulta, `IntForByes` guardará somente os dados apresentados no arquivo `codigo_consultando.PRG`. Para restaurar o quadro global de uma consulta prévia, `IntForBayes` utilizará estes dados e o algoritmo de propagação de grau de certeza (ver 4.2.3.2).

Na gravação do arquivo `codigo_consultando.PRG` só serão registradas as evidências que foram informadas e que possuam graus de certeza diferentes de zero. As evidências informadas com grau de certeza zero ("não sei"), retornarão, na continuação da consulta, ao status inicial ("não informado"), garantindo que possam ser perguntadas (possivelmente, agora, sabemos mais sobre elas).

Os arquivos utilizados por `IntForBayes` estão dispostos em diversos diretórios, conforme podem ser vistos na figura 4.3 a seguir.

consumido bastante tempo de desenvolvimento, tem seus algoritmos tão fácil e obviamente antevistos do mero entendimento do manual do usuário e da descrição das estruturas de dados, que não julgamos necessário nem apropriado descreve-los numa dissertação de mestrado. O oposto pode ser dito do módulo de inferências, porisso passamos a detalhar (em alto nível) seus principais algoritmos.

4.2.3 Módulo de Inferências

4.2.3.1 Módulo Principal de Inferências

O módulo de inferências de IntForBayes é utilizado no processamento de consultas às bases de conhecimento editadas pelo módulo de edição. Em alto nível de abstração, este módulo é apresentado na figura 4.4. Resaltamos apenas que quando escolhida uma base de conhecimento válida, serão transportados para a memória os arquivos com extensões ".RED", ".LSP", ".LSF", ".LGH", ".LGE", ".LSH" e ".LSE" do diretório particular da referida base (ver figura 4.3).

```

leia o nome das bases de conhecimento existentes
repita
| apresente os nomes das bases de conhecimento existentes
| obtenha o nome de uma base de conhecimento para consultar
| se o nome da base de conhecimento for igual a "Esc"
|   então encerre a fase de inferências
|   senão carregue a base de conhecimento escolhida
| repita
| | leia o identificador do consultando
| | se o identificador do consultando for igual a "Esc"
| |   então abandone a base de conhecimento escolhida
| |   senão
| |     se o consultando não for cadastrado
| |     então
| |       | pergunte se deseja cadastrá-lo
| |       | se desejar cadastrá-lo
| |       |   então
| |       |   | cadastre o consultando
| |       |   | processe uma consulta
| |       |   senão volte a ler o identificador do consultando
| |       senão processe uma consulta
| até que o identificador do consultando seja igual a "Esc"
até que o nome da base de conhecimento seja igual a "Esc"

```

Figura 4.4 - Algoritmo do Módulo Principal de Inferências

4.2.3.2 Processamento de uma Consulta

A figura 4.5 apresenta o algoritmo de processamento de uma consulta à IntForBayes.

Ressaltamos alguns detalhes que poderiam passar despercebidos:

- a) Os comentários dos consultandos estão no arquivo (do diretório CDNS) que tem nome igual ao código do consultando e extensão ".TXT", mostrado na figura 4.3)
- b) Cada consulta é armazenada no arquivo (do diretório da base de conhecimento escolhida) com nome igual ao código do consultando e com extensão ".PRG" (ver figura 4.3). Este arquivo armazena as informações estritamente necessárias para a restauração da última

consulta do consultando que foi armazenada.

c) As hipóteses de primeiro nível são apresentadas por ordem decrescente de suas probabilidades a priori, quando nenhum grau de certeza das evidências foi informado, e por ordem decrescente de seus graus de certeza, se algum grau de certeza de alguma evidência foi informado. Os graus de certeza mínimo e máximo são calculados, quando necessário, a partir dos valores dos campos Piso Alcançável e Teto Alcançável, respectivamente, e do campo Chance a Priori da tabela Rede, através da fórmula (8) de 2.3 trocando-se a chance atual por piso alcançável e por teto alcançável, respectivamente.

se o consultando for antigo

então

leia os comentários editados na última consulta

apresente os comentários para edição

edite comentários sobre o consultando

se o consultando possui alguma consulta gravada

então

pergunte se o usuário deseja reativar a última consulta gravada

se desejar reativar a última consulta gravada

então

para todas as evidências informadas na consulta gravada

propague o efeito de seus graus de certeza

repita

! informe a classificação global das hipóteses de primeiro nível

! apresente as opções disponíveis ("D", "Entra", "C", "R", "Esc")

! obtenha uma opção válida

! caso a opção seja

! "D" : Ofereça/corrija dados

! "Entra" : investigue a hipótese desejada

! "C" : mostre Como chegou ao resultado da hipótese desejada

! "R" : Altere o ponto de satisfação

! "Esc" : abandone o ciclo de opções disponíveis

até que a opção seja igual a "Esc"

se desejar atualizar os comentários sobre o consultando

então edite comentários

se desejar gravar a consulta atual

então grave a consulta atual

FIGURA 4.5 - Algoritmo de Processamento de uma Consulta

4.2.3.3 Propagação do Grau de Certeza de um Fato

O algoritmo de propagação de um grau de certeza (figura 4.6) deve ser utilizado sempre que houver alguma atribuição de um valor válido de grau de certeza a uma evidência. A sua finalidade é atualizar os campos das tabelas constituintes da base de conhecimento que dependem direta ou indiretamente do valor de grau de certeza informado. Estes campos são Grau de Certeza, Teto Alcançável, Piso Alcançável e Chance Atual, da tabela Rede; e FAuxChance, FAuxTeto e FAuxPiso, da tabela LstPais. As fórmulas utilizadas na atualização desses campos encontra-se em 2.2 e 2.3. O algoritmo é recursivo e o sentido da propagação é o da evidência para as hipóteses de primeiro nível, passando por todos os fatos intermediários que de alguma forma dependem desta evidência.

A faixa de valores utilizada pelo usuário no fornecimento do grau de certeza de uma evidência varia de -5 (descrença absoluta na evidência) a +5 (crença absoluta na evidência). A faixa de valores utilizada internamente por IntForBayes varia de -1 a +1. Assim, quando o usuário informa o grau de certeza de uma evidência, IntForBayes simplesmente guarda o valor digitado dividido por +5. Este procedimento garante uma maior simplificação numérica nas fórmulas utilizadas na propagação do grau de certeza. Na apresentação de valores de graus de certeza para o usuário, simplesmente multiplicamos estes valores por +5.

```

faça para todos os pais de fato
  se pai for diferente de raiz absoluta
    então
      calcule os novos fatores atualizadores da chance
        atual, teto alcançável e piso alcançável do pai
      atualize a chance atual, o teto alcançável e o piso
        alcançável do pai
      atualize o grau de certeza atual do pai
      guarde os fatores atualizadores da chance atual, teto
        alcançável e piso alcançável do pai
      propague o grau de certeza do pai (recursivamente)

```

FIGURA 4.6 - Algoritmo de Propagação de Grau de Certeza

4.2.3.4 Oferta/correção de Dados (Voluntariamento)

O algoritmo deste módulo encontra-se na figura 4.7.

Queremos ressaltar que:

a) Uma vez escolhida uma evidência o usuário poderá usar a opção "del" para simplesmente cancelar a informação prestada anteriormente para o grau de certeza (caso exista). Usando esta opção será utilizado o algoritmo de despropagação de graus de certeza. Este algoritmo é análogo ao de propagação. Sua função basicamente consiste na propagação de um grau de certeza zero da evidência (além do retorno do status do seu grau de certeza à situação original de "não perguntado") levando em conta o cálculo dos valores do teto alcançável e piso alcançável dos fatos influenciados pela evidência, para refletir a nova situação desta evidência.

b) Além da opção "del" o usuário poderá digitar o valor do grau de certeza que desejar para a evidência, na escala de -5 (descrença absoluta na evidência) a +5 (crença absoluta na evidência), independentemente do valor atual do grau de certeza da evidência. Com a digitação de um valor válido IntForBayes

transfere este valor para a faixa de -1 a +1 (basta dividir o valor por +5) e invoca o algoritmo de propagação de grau de certeza. Vale salientar que a propagação de um grau de certeza válido ao longo da base de conhecimento, devido às fórmulas a que chegamos para a propagação, já anula automática e eficientemente os efeitos anteriores do último valor informado do grau de certeza desta evidência. Portanto, no caso de alteração do grau de certeza de uma evidência, basta o usuário digitar o novo valor por cima do valor anterior (apresentado na tela): com a simples propagação do novo valor digitado ocorrerá simultaneamente a anulação dos efeitos anteriores do grau de certeza antigo e a atualização dos efeitos do novo grau de certeza.

```

repita
| apresente os grupos de evidências disponíveis
| escolha um grupo de evidência
| se o grupo escolhido for igual a "Esc"
|   então abandone este loop (externo)
|   senão
|     repita
|       apresente as evidências do grupo escolhido
|       escolha uma evidência
|       se a evidência escolhida for igual "Esc"
|       então abandone este loop (interno)
|       senão
|         leia e atribua-lhe um grau de certeza válido
|         se o grau de certeza atribuído for igual a "Del"
|         então despropague o grau de certeza anterior da
|                                     evidência
|         senão propague o grau de certeza atual da evidência
|     até que a evidência escolhida seja igual a "Esc"
até que o grupo escolhido seja igual a "Esc"

```

Figura 4.7 - Algoritmo de Oferta/correção de Dados (Voluntariamente)

4.3.2.5. Investigação de um fato

O algoritmo de investigação de um fato (figura 4.8) é usado quando o usuário deseja determinar o grau de certeza de uma hipótese de primeiro nível (opção "Entra" do algoritmo da figura 4.5). Ele é usado recursivamente na investigação dos fatos que influenciam direta ou indiretamente a hipótese de primeiro nível escolhida pelo usuário para uma investigação.

```
se a importância do fato é maior ou igual ao ponto de satisfação
então
  se o fato for perguntável
  então
    | informe a situação da hipótese de primeiro nível sendo
    |                                     investigada
    | obtenha o grau de certeza do fato
    | se o grau de certeza obtido for igual a "Esc"
    | então faça QuerAbandonarInvestigação igual a verdade
    | senão
    |   propague o grau de certeza do fato
    |   atualize a importância dos fatos pendentés
  senão
  enquanto a importância do fato for maior que o ponto de
  | satisfação e não QuerAbandonarInvestigação e tem
  |                                     filho não examinado
  | empilhe o fato em fatos pendentés
  | obtenha o filho mais importante do fato
  | se a importância do filho mais importante do fato
  |   for maior que o ponto de satisfação
  |   então investigue o filho mais importante do fato
  |                                     {recursivamente}
  fim
```

FIGURA 4.8 - Algoritmo de Investigação de um Fato

Em cada início da investigação de um fato, será verificado se a importância do fato é menor ou igual ao ponto de satisfação estabelecido. Em caso afirmativo a investigação do fato atual é finalizada. Caso contrário a investigação prosseguirá com a verificação do tipo de fato que está sendo investigado. Se o campo TagCar da tabela Rede (figura 4.1) for igual a "P" trata-se

de um fato perguntável e se for igual a "D" trata-se de um fato dedutível.

Se o fato for perguntável então serão informados na tela os valores dos graus de certeza atual, mínimo e máximo da hipótese de primeiro nível que está sendo investigada. Em seguida será solicitado do usuário qual o grau de certeza daquele fato perguntável. Caso o usuário responda com "Esc" então será dada por encerrada a fase de investigação da hipótese de primeiro nível. Caso o usuário responda à pergunta com um grau de certeza válido o seu efeito (nas chances e graus de certeza de todos os ascendentes, e nas importâncias dos ascendentes presentes no caminho entre a hipótese de primeiro nível e o fato sendo perguntado) será propagado ao longo da base de conhecimento. Além disso o usuário neste ponto pode perguntar ao sistema para que tal pergunta está sendo feita e neste caso IntForBayes justificará (nível por nível, enquanto desejado ou não se exaurir a sub-árvore) o objetivo da pergunta. Estes fatos encontram-se armazenados em uma pilha que representa o caminho percorrido na investigação da hipótese de primeiro nível. Feito isto será dada por encerrada a fase de investigação do fato perguntável.

Se o fato for dedutível haverá a entrada em um ciclo cujo abandono se dará quando for constatado que a importância do fato é menor ou igual ao ponto de satisfação estabelecido, ou quando o usuário exprimiu o desejo de abandonar a fase de investigação, ou ainda quando, no processo de escolha de outros fatos para investigação, for verificado que o atual filho mais importante do fato é igual ao do ciclo interior. Esta última condição foi

incluída pois na investigação de um fato poderia ocorrer uma situação em que sua importância é maior que o ponto de satisfação e a importância de todos os seus filhos menor. Neste caso, sem a última condição ocorreria um loop infinito.

No início do ciclo o fato (representado pelo seu índice na tabela Rede da figura 4.1) é colocado no topo da pilha que guarda todos os fatos que se encontram atualmente em fase de investigação. Uma das finalidades desta pilha é auxiliar o processo de justificção da pergunta escolhida, pois guarda a trilha de fatos pendentes da investigação da hipótese de primeiro nível. A outra finalidade é permitir a atualização das importâncias dos fatos sendo investigados (necessária no teste de continuidade de cada ciclo).

Em seguida é feita a escolha, entre os filhos do fato mais atual sendo investigado, do fato que possui maior valor de importância.

Caso o fato escolhido possua importância maior que o ponto de satisfação estabelecido então este fato passará a ser o novo fato a ser investigado.

Finalmente, após a investigação deste fato ou caso sua importância seja menor que o ponto de satisfação, será atualizada a pilha dos fatos pendentes, simplesmente retirando o fato que se encontra no topo da pilha.

5. TESTES DE DESEMPENHO (NÚMERO DE PERGUNTAS X PRECISÃO)

Os procedimentos asseguradores de qualidade usualmente usados em projetos semelhantes (isto é, revisão crítica do código por colegas, testes de caixa preta e caixa branca, testes funcionais e estruturais, testes dos módulos individuais, testes de fronteira e teste de integração), foram todos feitos e nada temos de relevante a relatar sobre eles. Temo-lo sobre os testes de desempenho.

Na primeira seção deste capítulo revisaremos o "background" e a evolução histórica de IntForBayes. Na segunda seção consideraremos os aspectos a serem avaliados no algoritmo. Na terceira seção apresentaremos as bases de conhecimento utilizadas no testes. Na quarta seção informaremos como os testes foram planejados. Na quinta e última seção mostraremos as conclusões obtidas com os testes.

5.1 EVOLUÇÃO HISTÓRICA DENTRO DO PROJETO INTFORBAYES

Em primeiro lugar, o trabalho de [NUNE 86] motivou o nosso. Naquele, o algoritmo estava detalhado apenas para redes bayesianas de dois níveis, e escolhia para perguntar a evidência mais importante, segundo a fórmula:

$$\text{Importância}(E) = \sum_H (FS(E,H) * C(H) - FN(E,H) * C(H)) \quad (1)$$

- onde:
- O somatório é feito sobre todas as hipóteses pais de E;
 - O significado dos demais símbolos encontra-se em 2.2.

Em segundo lugar, notamos que o trabalho de [NUNE 86] podia ser melhorado de vários modos:

- 1) Não limitar o número de níveis da rede de conhecimento a 2;
- 2) Permitir que o usuário ofereça informações voluntárias a qualquer ponto de uma consulta;
- 3) Possibilitar que os critérios automáticos de poda sejam facilmente ajustáveis de acordo com as precisões exigidas nas aplicações; e
- 4) Apresentação dos graus de certeza extremos alcançáveis pelas hipóteses sendo investigadas, aliadas com a possibilidade de interferência por parte do usuário na parada do processo.

Em particular, observamos que a fórmula original para cálculo da importância de uma evidência tinha os seguintes pontos melhoráveis:

- 1) Transformação dos valores obtidos pela fórmula (de 0.0 até o maior real da máquina) em uma escala mais clara e conveniente para compreensão dos usuários (de 0.0 a 1.0, por exemplo), e
- 2) Maior justiça na escolha de fatos que contribuam decisiva ou substancialmente para a negação da hipótese investigada. Notemos que os termos internos ao somatório correspondem à diferença entre os fatores do sim e os fatores do não de cada arco (filho-pai), multiplicados pela chance atual do pai. Os valores práticos dos fatores do sim são quase sempre elevados (da ordem de $(10 \text{ elevado a } +1)$ a $(10 \text{ elevado a } +6)$ em SINDROMUS [NICO 87], por exemplo) e fatores do não em geral muito pequenos (da ordem de $(10 \text{ elevado a } -1)$ a $(10 \text{ elevado a } -4)$, no mesmo S.E.). A diferença entre estes fatores será

praticamente igual ao fator do sim. Desta forma os filhos cujas ligações com as hipóteses em análise apresentam fatores do não muito próximos de zero (que podem ser decisivos para a negação da hipótese) não são valorizados justamente.

Assim, propusemos a seguinte fórmula:

$$\text{Imp}(E) = \text{Imp}(E, H_1, \dots, H_n) \quad (2)$$

sendo:

- n o número de hipóteses relacionadas diretamente com E;

$$\text{Imp}(E, H_1, \dots, H_k) = \text{Imp}(E, H_1, \dots, H_{k-1}) + \text{Imp}(E, H_k) * [1 - \text{Imp}(E, H_1, \dots, H_{k-1})];$$

com:

$$\text{Imp}(E, H_k) = P(H_k : E.teto) - P(H_k : E.piso).$$

O artigo [LYRA 89] apresenta maiores detalhes.

Em terceiro lugar, raciocinamos que, no uso típico dos S.E.'s diagnosticadores (subconjunto mais popular dos classificadores), o cálculo da importância podia e devia se restringir ao espaço de pesquisa sob a hipótese de primeiro nível escolhida, ao invés de ser feito globalmente.

Assim, modificamos a fórmula para a sua versão "local":

$$\text{Imp}(F, H) = P(H : F.teto) - P(H : F.piso) \quad (3)$$

onde:

- $\text{Imp}(F, H)$ é a importância do fato (F) para hipótese (H);
- $P(H : F.teto)$ é a probabilidade alcançada por H se F atinge o teto; e
- $P(H : F.piso)$ é a probabilidade alcançada por H se F atinge o piso.

Nossa conjectura (de ser melhor localizar do que globalizar o cálculo das importâncias) foi confirmada por testes descritos

adiante.

Notamos também que vários conceitos do trabalho de [NUNE 86] e [FERR 90] tornavam os algoritmos mais complexos (não tínhamos certeza de que funcionavam bem em todos os casos possíveis) e eram dispensáveis. Os conceitos de:

- importância de um fato
- limiar de aceitação (um para cada hipótese)
- limiar de rejeição (um para cada hipótese)
- hipótese campeã por antecipação
- hipótese classificada
- hipótese em disputa
- hipótese desclassificada
- ...

podem ser substituídos apenas pelos conceitos de:

- importância de um fato e
- ponto de satisfação

Em quarto lugar, uma vez que:

- a) perguntas, cálculos intermediários e resultados usam todos graus de certeza;
- b) graus de certeza não se relacionam linearmente com probabilidades,

então deveria ser melhor calcularmos importâncias em termos de graus de certeza. Assim, propusemos a fórmula:

$$\text{Imp}(F,H) = (GC(H:F.teto) - GC(H:F.piso))/2 \quad (4)$$

onde :

- $\text{Imp}(F,H)$ é a importância do fato (F) para a hipótese (H);

- $GC(H:F.teto)$ é o grau de certeza alcançado por H se F atingir o teto;
- $GC(H:F.piso)$ é o grau de certeza alcançado por H se F atingir o piso

A vantagem desta fórmula foi confirmada por testes adiante descritos. O denominador 2 foi usado para mapear a importância para o intervalo $[0,+1]$, mais natural na definição do ponto de satisfação.

5.2 ASPECTOS A SEREM AVALIADOS NO ALGORITMO

No processo evolutivo que resultou no atual algoritmo de IntForBayes, foram tomadas certas decisões visando uma melhor performance global. Essas decisões foram tomadas após as ponderarmos muito. No entanto, por mais seguras ou mesmo óbvias que elas nos pareceram, não foram justificadas por prova formal nem por testes conclusivos. Aqui, queremos avaliar os efeitos dessas decisões.

Quatro aspectos do algoritmo de IntForBayes são de vantagens menos óbvias-seguras e de avaliação quantitativa mais fácil através de testes simples, mesmo que extensos:

- a) qual a vantagem de fazermos o cálculo da importância localmente sobre o fazermos globalmente;
- b) qual a vantagem, para o cálculo da importância, de usarmos a fórmula 4 (baseada em graus de certeza) sobre usarmos a fórmula 3 (baseada em probabilidades);

- c) qual a vantagem de fazermos sempre a pergunta mais inteligente, mais importante, sobre fazermos as perguntas em uma ordem cega; e
- d) qual a vantagem de pararmos a determinação do grau de certeza de uma hipótese assim que sua importância descer ao ponto de satisfação, sobre a pararmos somente por exaustão.

Dois outros aspectos de IntForBayes são de vantagens mais óbvias porém de avaliação quantitativa mais difícil, exigindo construção de muitos grandes S.E.'s reais a serem testados em muitos casos reais:

- e) qual a vantagem média, em problemas práticos, de permitirmos a oferta voluntária de algumas informações obtidas com custo desprezível; e
- f) qual a vantagem média, em problemas práticos, de procedermos por níveis de exigência, conforme aconselhado em 3.3.

5.3 DUAS BASES DE CONHECIMENTO USADAS NOS TESTES

A primeira base de conhecimento bayesiano escolhida foi a de SINDROMUS [NICO 87], que é um S.E. bayesiano para o diagnóstico de síndromes de malformações congênitas. Até onde conhecemos, ele é o maior S.E. bayesiano desenvolvido no País, com 260 hipóteses (todas elas de primeiro nível), 700 evidências (agrupadas em 57 grupos), cada pai tendo em média 14 filhos, e cada filho tendo em média 5 pais. Vide a figura 5.1. SINDROMUS foi escolhido por seu grande tamanho; por ter sido testado em casos reais (na Escola Paulista de Medicina), obtendo um índice

de acertos considerado excelente; por ser o único sistema bayesiano de tal porte que nos era disponível; por ter sido desenvolvido localmente; e por ser, com sua simples estrutura de dois níveis, representante de uma ampla classe de S.E.'s.

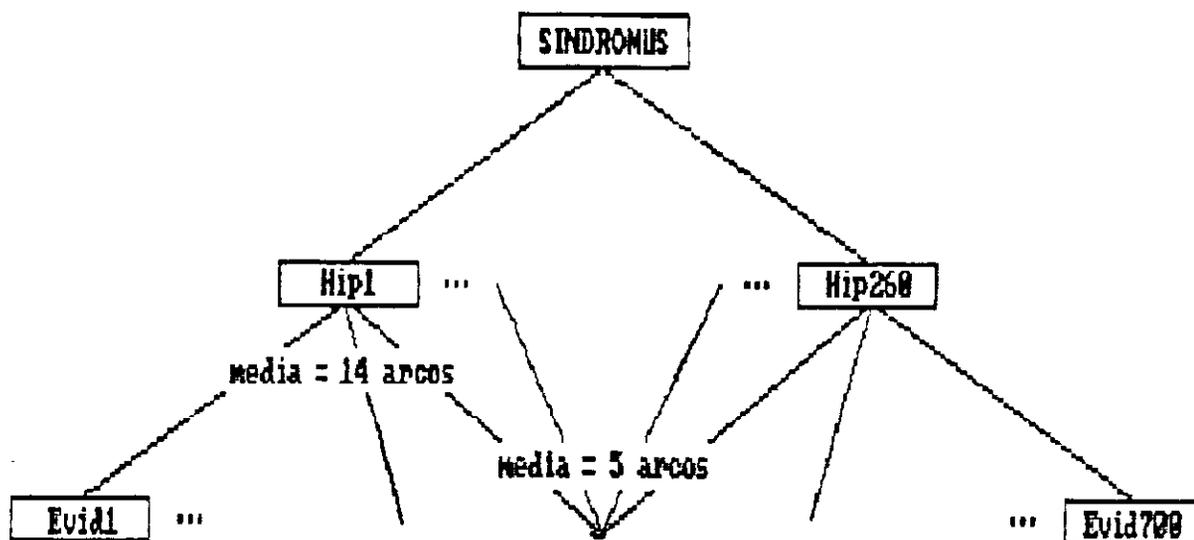


Figura 5.1- Esboço da Estrutura de SINDROMUS.

Buscamos outros S.E.'s bayesianos de estrutura o mais diferente possível de SINDROMUS (isto é, com estrutura estreita e profunda), mas não conseguimos ter acesso suficiente a nenhum deles. Prospector [GASC 81] tem apenas uma pequena parte da sua base de conhecimento (sobre Geologia) disponível a todos. Idem quanto ao sistema diagnosticador de doenças coronarianas desenvolvido por Machado [MACH 85]. Assim, premidos pelo tempo, criamos uma pequena rede bayesiana, com as características desejadas, adequada aos nossos propósitos, mas sem corresponder ao mundo real. Os valores das probabilidades foram fixados de forma a refletirem situações típicas encontradas no SINDROMUS e no Prospector. Vide a figura 5.2.

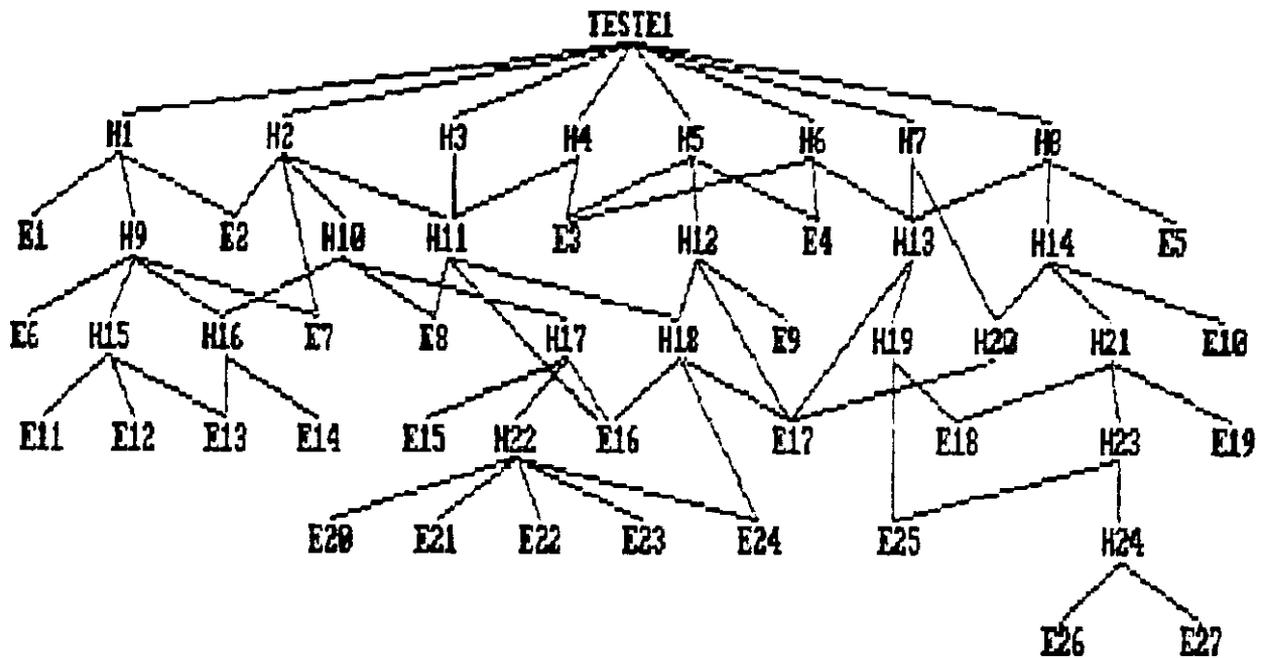


Figura 5.2 - A Estrutura do S.E. Teste1

5.4 PLANEJAMENTO DOS TESTES

Para avaliarmos o aspecto "a" de 5.2 (vantagem do cálculo local sobre o cálculo global das importâncias), fizemos 2 versões de IntForBayes, a serem comparadas:

IntGlobProb - escolhe a pergunta mais Inteligente, mais importante;

- calcula as importâncias Globalmente;
- a fórmula da importância baseia-se em Probabilidades; e

IntLocProb - idem;

- calcula a importância Localmente;
- idem;

Para avaliarmos o aspecto "b" de 5.2 (vantagem da importância basear-se em graus de certeza, ao invés de probabilidades), fizemos uma terceira versão de IntForBayes, a ser comparada com IntLocProb:

IntLocGC - escolhe a pergunta mais Inteligente, mais importante;

- calcula as importâncias Localmente; e
- a fórmula da importância baseia-se em GC's

Para avaliarmos o aspecto "c" de 5.2 (vantagem de fazermos sempre a pergunta mais importante sobre fazermos às cegas), fizemos uma quarta versão de IntForBayes, a ser comparada com IntLocGC:

CegLocGC - escolhe as perguntas às Cegas;

- calcula as importâncias Localmente;
- a fórmula da importância baseia-se em GC's.

Para avaliarmos o aspecto "d" de 5.2 (vantagem de pararmos por alcançar o ponto de satisfação, ao invés de pararmos somente por exaustão), percebemos que necessitaríamos rodar os mesmos casos de consultandos para pontos de satisfação variando de 0.0 (o que equivale a fazermos todas as perguntas sob a hipótese de primeiro nível escolhida e só pararmos por exaustão) a 1.0 (o que equivale a uma parada antecipada, sem fazer nenhuma pergunta).

Planejamos os nossos testes segundo o seguinte procedimento:

```

para rede em {SINDROMUS, Teste1}:
| crie bateria de casos {SINDROMUS=100 casos; Teste1=125 casos}
| para cada caso
| | voluntarie todas as evidências e considere os resultados como
| | | | | "a verdade"
| fim
| para versão-de-IntForBayes em {IntGlobProb, IntLocProb,
| | | | | IntLocGC, CegLocGC}
| | para ponto-de-satisfação variando de 0.0 a 1.0
| | | para cada caso
| | | | para todas as hipóteses
| | | | | escolha uma hipótese
| | | | | rode versão-de-IntForBayes
| | | | | compare GC da hipótese com "a verdade"
| | | | | acumule estatísticas
| | | | | fim
| | | fim
| | | imprima estatísticas
| | fim
| fim
| cobrindo tanto SINDROMUS como Teste1, compare os resultados de:
| | IntGlobProb x IntLocProb
| | IntLocProb x IntLocGC
| | IntLocGC x CegLocGC
|
fim

```

onde um caso, armazenado em uma tabela, é uma atribuição de um GC (gerado aleatoriamente no intervalo -5 a +5) a cada uma das evidências de uma rede. Quando a versão de IntForBayes, modificada para os testes, necessita do GC de uma evidência, ao invés de perguntá-lo ao usuário, ela o obtém da tabela armazenadora do caso.

5.5 CONCLUSOES SOBRE OS TESTES

Os resultados dos testes realizados encontram-se no apêndice B. Algumas explicações são úteis à leitura desses relatórios: para cada teste foram criados aleatoriamente diversos casos de consultas através da geração e atribuição de graus de

certeza a todas às evidências. Para cada caso foram obtidos e armazenados os resultados de todas as hipóteses de primeiro nível deduzidos a partir dos graus de certeza atribuídos a todas as evidências. Estes resultados obtidos foram considerados como "a verdade" para efeito de comparação. Em seguida solicitamos a investigação das hipóteses de primeiro nível, uma em cada consulta (para pontos de satisfação variando de 0.0 a 1.0) e comparamos os resultados de cada consulta com "a verdade" esperada para a hipótese escolhida. Para a rede Teste1 (figura 5.1) foram criados aleatoriamente 125 casos de consultas e para cada caso foram investigadas as 8 hipóteses de primeiro nível, perfazendo um total de 1000 situações diferentes. Para a rede SINDROMUS (figura 5.2) foram escolhidas aleatoriamente 130 hipóteses de primeiro nível (50% do total), criados aleatoriamente 100 casos de consultas e para cada caso foram investigadas as 130 hipóteses escolhidas, perfazendo um total de 13.000 situações diferentes.

Uma cuidadosa análise comparativa dos resultados dos testes realizados, permite-nos chegar às seguintes conclusões (usaremos as mesmas letras dos itens de 5.2):

a) Com a rede Teste1, observamos que:

- 1) fazendo-se, por exemplo, 58% das perguntas, IntLocProb deu melhor resultado (desvio padrão menor, 0.129) do que IntGlobProb (0.143); outra interpretação é:
- 2) para uma mesma precisão (desvio padrão 0.129), IntLocProb deu um resultado melhor (menor número de perguntas, 58%) do que IntGlobProb (62%, aproximadamente). Similar observação vale

para a rede SINDROMUS.

Concluimos que, em S.E.'s que tendem mais para diagnosticadores que para classificadores, o cálculo localizado da importância é algo melhor do que globalizado. Isto já era esperado intuitivamente, conforme mencionamos em 5.1

b) Com a rede Teste1, observamos que:

- 1) fazendo-se, por exemplo, 60% das perguntas, IntLocGC deu um melhor resultado (desvio padrão menor, 0.036) do que IntLocProb (0.067); outra interpretação é:
- 2) para uma mesma precisão (por exemplo, desvio padrão 0.036), IntLocGC deu um resultado melhor (menor número de perguntas, 60%) do que IntLocProb (80%, aproximadamente).

Similar observação vale para a rede SINDROMUS.

Concluimos que a fórmula baseado em GC's, para o cálculo das importâncias, é sensivelmente melhor que a fórmula baseada em probabilidades. Isto confirma nossa conjectura e argumentação em 5.1.

c) Com a rede Teste1, observamos que:

- 1) fazendo-se, por exemplo, 60% das perguntas, IntLocGC deu um melhor resultado (desvio padrão menor, 0.036) do que CegLocGC (0.25); outra interpretação é:
- 2) para uma mesma precisão (por exemplo, desvio padrão 0.036), IntLocGC deu um resultado melhor (menor número de perguntas, 60%) do que CegLocGC (90%, aproximadamente).

Similar observação vale para a rede SINDROMUS, só que com valores mais impressionantes.

Concluimos que fazermos sempre a pergunta mais inteligente,

mais importante, é bastante melhor que fazemos as perguntas às cegas. Isto já era esperado, pois em muitas outras aplicações de I.R. o uso de heurísticas tem demonstrado vantagens sobre buscas cegas.

- d) Na rede Teste1 observamos que, usando-se IntLocGC com um certo ponto de satisfação (por exemplo 0.20), chegamos ao resultado com um número de perguntas (44%) bem menor do que se procedermos por exaustão (número de perguntas = 100%, que equivale ao ponto de satisfação 0.00), desde que assumamos como aceitável a precisão do desvio padrão 0.10; Similar observação vale para a rede SINDROMUS. Tudo isto era esperado, intuitivamente.

Concluimos pela substancial vantagem de utilizarmos o conceito de ponto de satisfação, ao invés de somente pararmos por exaustão.

- d') Comparando-se os resultados das aplicações de IntLocGC às redes Teste1 e SINDROMUS observamos que, para obtermos uma mesma precisão (por exemplo, desvio padrão 0.10), os pontos de satisfação variam muito de uma rede para outra (0.20 para Teste1, 0.001 para SINDROMUS), mesmo que o número de perguntas seja praticamente igual (44% em ambos os casos).

Concluimos que o "default", fixado na edição, para o ponto de satisfação, deve ser escolhido diferentemente para cada rede (para Teste1, escolheríamos 0.20, para SINDROMUS, 0.001). Não havíamos antevisto isto, inicialmente, mas devíamos tê-lo feito, ou ao menos suspeitado da possibilidade desses resultados, pois agora nos parecem bastante previsíveis.

Adotemos fixar o valor do ponto de satisfação em três níveis: o primeiro na configuração de IntForBayes, prevalecendo se não for redefinido na edição; o segundo na [re]edição de uma determinada rede de conhecimento, em função da experiência adquirida no uso do S.E., pesando-se os prós e os contras do conflito entre número de perguntas (custo) e precisão dos resultados (benefício); o terceiro, válido só durante uma sessão e perdido ao seu final.

e+f) São significativos os ganhos com a adoção da localidade e graus de certeza no cálculo das importâncias, com fazermos sempre a pergunta mais importante, e com trabalharmos com um ponto de satisfação. Porém (embora não tenhamos podido quantificá-los através do uso real de muitos S.E.'s reais), muito, muito maiores são os ganhos resultantes da combinação dos enfoques supra-citados com a adoção de voluntariamento e o procedermos por níveis de exigências, conforme aconselhado em 3.3. Por exemplo, com a rede de SINDROMUS, o enfoque IntLocGC poderia exigir tipicamente 2 voluntariamentos e 2 respostas a perguntas, com certo ponto de satisfação, para chegar a um certo desvio padrão. O motor de inferências original de SINDROMUS (com voluntariamento mas com um conceito mais simples de importância, classificando as evidências, estaticamente, em Diferenciais, Frequentes e Ocasionais, com FS e FN iguais para todas as evidências de uma mesma classe) poderia exigir 2 voluntariamentos, e 3 respostas a perguntas para chegar à mesma precisão, cabendo ao usuário abortar o processo de inferência. Um sistema, que exaurisse o sub-grafo sob a hipótese de primeiro nível escolhida necessitaria uns 14

(média do número de filhos de uma hipótese)
voluntariamentos/respostas. Um sistema que funcionasse
exaurindo todo o grafo exigiria 700
voluntariamentos/respostas!

6. CONCLUSOES E SUGESTOES

6.1 CONCLUSOES

A utilização de IntForBayes na construção e consultas aos dnis S.E.'s usados como teste permitiram-nos concluir que, apesar do "shell" estar longe de ter alcançado sua forma evolutiva final, já alcançou seus objetivos iniciais (enunciados em 1.2), apresenta vantagens e mostra-se promissor:

- 1) O editor de conhecimento, apesar da sua simplicidade e de não ser decisivo nesta dissertação, já é plenamente satisfatório tanto em facilidade/comodidade (ver 3.2.2.1) como em eficiência tempo-espaço.
- 2) A escolha da base (frequentemente feita uma única vez por sessão), a identificação e/ou cadastramento dos consultandos (por função de hash), a edição de comentários, a escolha de hipóteses, e a apresentação dos resultados, todos têm seus tempos de C.P.U. negligenciáveis.
- 3) No voluntariamento, a reunião das evidências em grupos é cômoda para o usuário. E a propagação (tempo imperceptível) dos efeitos de um grau de certeza tão logo ele seja fornecido, ao invés de propagá-los todos ao final do voluntariamento (tempo possivelmente perceptível) tem vantagens óbvias.
- 4) Gravar somente as evidências informadas, no final de uma consulta, e, a partir daí, numa reconsulta, calcular os graus de certeza de todos os fatos afetados, revelou-se não só mais econômico em espaço como em tempo (pois as leituras

adicionais, em discos rígidos usuais, demoram mais que estes cálculos adicionais).

- 5) Na propagação de graus de certeza as fórmulas (9 e 12), que introduzimos em 2.2 e em 2.3, são bem mais eficientes que as fórmulas tradicionais.
- 6) Quanto ao número de perguntas e à precisão de resultados (que são de longe as duas grandes medidas de eficiência de um S.E.), podemos afirmar que, para os S.E.'s teste, que tendem mais para diagnosticadores do que para classificadores:
 - i) O cálculo localizado de importância (que adotamos), para um mesmo número de perguntas deu precisão cerca de 10% maior que o globalizado, e para uma mesma precisão requereu cerca de 6% menos perguntas.
 - ii) O cálculo que adotamos para a importância, baseado em graus de certeza e não em probabilidades, resultou cerca de 46% mais preciso, para um certo número de perguntas, e requereu cerca de 25% menos perguntas para uma certa precisão.
 - iii) Fazermos sempre a pergunta mais inteligente (ou seja, mais importante) no sentido de provar ou desprovar mais rapidamente a hipótese de primeiro nível escolhida, em comparação com fazermos as perguntas numa ordem cega (quer estática quer dinamicamente determinada), resultou cerca de 85% mais preciso, para um certo número de perguntas, e requereu cerca de 33% menos perguntas para uma certa precisão.
 - iv) A parada do processo de investigação de uma hipótese antecipada para tão logo a importância da mesma desça até um ponto de satisfação, resultou numa redução de 66% em relação ao procedimento por exaustão, para uma determinada precisão.

v) Redes diferentes apresentam pontos de satisfação ideais diferentes. O valor "default" para o ponto de satisfação, fixado na edição, deve ser escolhido diferentemente para cada rede de conhecimento. Essa escolha é feita por tentativas, podendo ser feita à semelhança do método da bisseção usado em Cálculo Numérico. Nas redes Teste1 e SINDROMUS optaríamos pela escolha inicial dos pontos de satisfação, iguais a 0.20 e 0.001, respectivamente.

vi) São expressivos os ganhos individuais sumariados em (i) a (iv). Muito maiores são os ganhos resultantes da combinação desses enfoques com o de voluntariamento e com o procedermos por níveis de exigências (ver 3.3). Por exemplo, para a rede de SINDROMUS [NICD 87] IntForBayes poderia precisar tipicamente de apenas 2 voluntariamentos e 2 respostas às perguntas; um "shell backward", mesmo focalizado para a mesma hipótese, poderia exigir cerca de 14 fornecimentos de informações; e um "shell" puramente "forward" poderia exigir até 700 informações.

6.2 SUGESTOES PARA TRABALHOS FUTUROS

Nossas sugestões são, na ordem em que gostaríamos de vê-las sendo postas em prática

- 1) Dar um melhor tratamento estatísticos na comparação dos resultados dos testes já feitos.
- 2) Fazer uma comparação entre os 4 enfoques (IntGlobProb, IntLogProb, IntLocGC e CegLocGG), mas agora para S.E.'s

tendendo mais para classificadores do que para diagnosticadores. Isto envolverá comparar um conjunto de pares (hipótese, grau de certeza) com outro conjunto similar.

- 3) Implementar os módulos de explanação "Como?" e "Para que?". Estes módulos deverão dar as explicações nível por nível, só passando de um para o próximo se for desejado pelo usuário.
- 4) Permitir que o engenheiro de conhecimento e o perito, ao criarem um S.E., definam campos para os comentários sobre os consultandos (figura 3.15), como, por exemplo, os seguintes campos:

NOME: String de até 30 caracteres (alfa-numéricos)

TELEFONE: Código: inteiro positivo, 3 dígitos

Número: inteiro positivo, 7 dígitos

...

OBS: String de até 128 caracteres (imprimíveis ou de controle de impressão)

- 5) Conceber e implementar um módulo de aprendizagem através de exemplos. Os exemplos serão divididos aleatoriamente em dois conjuntos, um para treinamento e outro para testes. Com o primeiro, o sistema gerará primeiramente uma rede, sem valores de GC, FS, FN, Cp, etc. O engenheiro de conhecimento e o perito corrigirão a estrutura desta rede, provavelmente introduzindo nodos intermediários (por exemplo, MAMIFERO, generalizando CARNIVORO, HERBIVORO, etc). Eliminarão também, tanto na rede como nos exemplos, os [quase] sinônimos e os [quase] antônimos. A seguir, o sistema extrairá o conhecimento probabilístico dos exemplos de treinamento. Esta extração será simples (meras estatísticas), precisa e sem grandes problemas

para tratar os "casos raros/com ruído/..." (bem discutidos por [GOME 89], no contexto de regras). Por fim, o engenheiro e o perito tomarão a rede gerada, a analisarão e a aplicarão aos exemplos do conjunto de teste, provavelmente descobrindo alguns pontos a serem melhorados manualmente.

- 6) Conceber e implementar um módulo que, com base na experiência por longo tempo com um S.E., sugira modificações nos valores numéricos de Chance a Priori, $P(E:H)$, $P(E:\sim H)$, Ponto de Satisfação, etc. ou mesmo na estrutura do sistema. Estas sugestões, se aprovadas pelos perito e engenheiro de conhecimento, com suas senhas, serão incorporadas ao sistema, guardando-se a versão anterior como "backup".
- 7) Usar IntForBayes para construir e testar em campo o maior número possível de S.E.'s reais, dos tipos mais diversos possíveis, chegando a resultados mais definitivos que os discutidos no cap 5. Somente com esta experiência poderão ser feitos importantes aprimoramentos tanto no campo das interfaces (especialmente de edição, explicações e apresentação de resultados), como nos próprios mecanismos internos de inferência, etc.
- 8) A partir de IntForBayes, conceber e implementar o módulo de execução de LIDIA [FERR 90], que é uma linguagem para S.E.'s, provida de variáveis, expressões aritméticas, operador genérico de inferências (que pode deslizar continuamente para simular AND-forte, AND-atenuado, OR-forte,...), procedimentos externos, etc. Ou, a partir de LIDIA, introduzir essas desejáveis características no "shell" IntForBayes.
- 9) Introduzir um tratamento dos correlacionamentos estatísticos

entre os fatos, à semelhança dos clusters/aglomerados de [MACH 85], [NORU 75a] e [NORU 75b], calcados nos testes do Chi-quadrado.

- 10) Permitir aos usuários que eles definam suas próprias funções de cálculo de importâncias, se o desejarem.
- 11) Idem a (7), porém após os aperfeiçoamentos de (8) a (10).
- 12) Dotar o editor de uma opção para visualizar graficamente a rede, à semelhança de VP-Expert, com "zoom-in", "zoom-out", etc.
- 13) Melhorar o editor, inspirando-se nas características dos editores de texto consagrados (onde for aplicável). Por exemplo, o editor poderia fazer operações de copiar/mover/eliminar/... blocos: blocos de evidências, blocos de grupos de evidências, blocos de hipóteses, etc.
- 14) Conceber e implementar um editor gráfico, no qual criaríamos a rede com um "mouse" e outros auxílios usuais em CAD e ICAD, visualizando-a constantemente na tela, com "zoom-in", "zoom-out", etc.

REFERENCIAS BIBLIOGRAFICAS

- [BHAT 90] Bhatnagar, R. K. and Kanal, L. N.
Handling Uncertain Information: A Review of Numeric
and Non-numeric Methods, in
Uncertainty in Artificial Intelligence (eds Kanal, L.
N. and Lemmer, J. F.), North - Holland, 3-26, 1990
- [BUCH 78] Buchanan, B. G. and Feigenbaum, E. A.
Dendral and Meta-Dendral: their applications dimension
Artificial Intelligence, volume 11, 1978
- [DUDA 79] Duda, R. O.
The Knowledge Acquisition System
Artificial Intelligence Center, SRI International,
Technical Report, 1979
- [ELST 79] Elstein, A. S., Shulman, L. S. and Sprafka, S. A.
Medical Problem Solving. An Analysis of Clinical
Reasoning
Harward University Press, 1979
- [FERR 90] Ferreira, J. U.
LIDIA - uma Linguagem para Desenvolvimento de S.E.'s com
Tratamento de Incertezas
Dissertação de mestrado, Departamento de Sistemas e
Computação, Universidade Federal da Paraíba, 1990

- [FORG 80] Forgy, C. L. and McDermott, J.
The DP55 User's Manual
Carnegie - Mellon University, Department of Computer
Science, Technical Report, 1980
- [FORS 84] Forsyth, R.
Expert Systems : Principles and Case Studies
Chapman and Hall, 1984.
- [GASC 81] Gaschnig, J.
PROSPECTOR: an expert system for mineral exploration
Infotech State of the Art, Report 9, N. 3, 1981
- [GOME 89] Gomes, F. A. C.
APREND: Um Sistema de Aquisição Automática de
Conhecimento a Partir de Exemplos
Dissertação de mestrado, Departamento de Sistemas e
Computação, Universidade Federal da Paraíba, 1989
- [GORR 68] Gorry, G. A., Barnett, G. O.
Experience with a model of sequential diagnosis
Computers and Biomedical Research, 490-507, 1968
- [HARM 88] Harmon, Paul et alii.
Expert Systems Tools and Applications
John Wiley & Sons, 1988
- [JENS 74] Jensen, K. and Wirth, N.
Pascal User Manual and Report
Springer, 1974

- [KASS 78] Kassirer, J. P. e Gorry, A.
Clinical Problem Solving: A Behaviourial Analysis
Ann. Int. Med., volume 89, 245-255, 1978
- [KUIT 88] Kuitca, A. M.
Resolucion de Problemas com Inteligência Artificial
EBAI (Escola Brasileiro-Argentina de Informática), 1988
- [LYRA 89] Lyra, I. A.
Núcleo de IntForBayes, uma Ferramenta para S.E.'s
Bayesianos
Anais do sexto Simpósio Brasileiro em Inteligência
Artificial, 195-205, 1989
- [MACH 85] Machado, R. J.
Sistema de Suporte à Decisão Médica - Diagnóstico
Diferencial e Geração de Planos de Investigação
Tese de D.Sc. da COPPE - UFRJ, 1985
- [MELL 80] van Melle, W.
A domain-independent system that aids in constructing
knowledge-based consultation programs
Doctoral Dissertation, Computer Science Department
Stanford University, 1980
- [NICO 87] Nicolletti, P. S.
Sindromus: Um Sistema Especialista para Diagnóstico de
Síndromes de Malformações Congênitas
Dissertação de Mestrado, Departamento de Sistemas e
Computação, Universidade Federal da Paraíba, 1987

- [NILS 71] Nilsson, N.
Problem Solving Methods in Artificial Intelligence
McGraw-Hill, 1971
- [NORU 75a] Norusis, M. J., and Jacques, J. R.
Diagnosis I. Symptom Nonindependence in Mathematical
Models for Diagnosis
Computers and Biomedical Research, 8, 156-172, 1975
- [NORU 75b] Norusis, M. J., and Jacques, J. R.
Diagnosis II. Diagnostic Models Based on Attribute
Clusters, a Proposal and Comparisons
Computers and Biomedical Research, 8, 173-188, 1975
- [NUNE 86] Nunes, M. G. V., Costa Pereira, R. E. e Monard, M. C.
O núcleo de um Sistema Especialista com Raciocínio
"Forward Chaining" Usando Inferência
Probabilística Bayesiana
Anais do XIII SEMISH, VI Congresso da Sociedade
Brasileira de Computação, 542-550, 1986
- [QUIN 82] Quinlan, J. R.
INFERNO: A Cautious Approach to Uncertain Inference
Rand Note N - 1898 - RC, 1982
- [ROTH 89] Rothenberg, J.
Expert System Tool Evaluation, in
Topics in Expert System Design: Methodologies and
Tools (eds Guida, G. and Tasso, C.), University of
Udine, 1989

[SHAF 76] Shafer, G.

A Mathematical Theory of Evidence
Princeton University Press, 1976.

[SHOR 78] Shortliffe, E. H.

Computer-Based Medical Consultation : MYCIN
North Holland Elsevier, 1978

[ZADE 78] Zadeh, L. A.

Fuzzy Sets as A Basis for A Theory of Possibility
Fuzzy Sets and Systems, volume 1, No. 1, 1978

APENDICE A

CONSTANTES DE CONFIGURAÇÃO DE INTFORBAYES

const

LimLstPF	= 4000	{número máximo de entradas em cada uma das tabelas LstPais e LstFilhos}
LimRede	= 2000	{Número máximo de fatos na rede de conhecimento}
LimLstGpo	= 200	{número máximo de grupos em cada uma das tabelas de Grupo de Hipóteses e Grupo de Evidências}
LimLstHips	= 800	{número máximo de hipóteses na tabela de hipóteses (LstHips)}
LimLstEvids	= 1400	{número máximo de evidências na tabela de evidências (LstEvids)}
LimFilhosDeUmPai	= 800	{número máximo de fatos ligados a um pai comum (é limite de array auxiliar usado na edição)}
LimBases	= 50	{número máximo de bases de conhecimento na tabela Bases}
LimNiveis	= 30	{número máximo de níveis de uma base de conhecimento (é limite de array auxiliar usado no cálculo das importâncias e nas explicações)}
TamTabCons	= 1001	{número inicial de registros da tabela de consultandos (a área de overflow fica depois deles)}

APENDICE B

RESULTADOS DOS TESTES

Ponto de Satisfação	Desvio Padrão	Número Médio de Perguntas
0.0000	0.0000	6.49(100.00%)
0.0100	0.0586	4.47(68.99%)
0.0200	0.1126	4.36(67.24%)
0.0400	0.1432	3.81(58.73%)
0.0600	0.1747	3.47(53.58%)
0.0800	0.1991	3.27(50.35%)
0.1000	0.2248	3.18(49.03%)
0.1500	0.2316	3.12(48.09%)
0.2000	0.2587	2.92(45.05%)
0.2500	0.2429	2.89(44.59%)
0.3000	0.2917	2.69(41.41%)
0.4000	0.3422	2.48(38.25%)
0.5000	0.5774	1.34(20.71%)
0.6000	0.5926	1.13(17.44%)
0.7000	0.6758	0.68(10.45%)
0.8000	0.6754	0.67(10.25%)
0.9000	0.6898	0.58(8.97%)

FIGURA B.1 - Resultados de IntGlobProb aplicado a Teste1

Ponto de Satisfação	Desvio Padrão	Número Médio de Perguntas
0.0000	0.0000	6.48(100.00%)
0.0100	0.0670	3.93(60.68%)
0.0200	0.1290	3.78(58.28%)
0.0400	0.1669	3.29(50.74%)
0.0600	0.1918	3.02(46.58%)
0.0800	0.2105	2.74(42.19%)
0.1000	0.2886	2.28(35.13%)
0.1500	0.3474	1.99(30.62%)
0.2000	0.4448	1.76(27.17%)
0.2500	0.4748	1.69(26.03%)
0.3000	0.6536	0.78(12.00%)
0.4000	0.6634	0.71(10.97%)
0.5000	0.6716	0.46(7.07%)
0.6000	0.7016	0.29(4.44%)
0.7000	0.7220	0.13(1.93%)
0.8000	0.7291	0.13(1.93%)
0.9000	0.7238	0.13(1.93%)

FIGURA B.2 - Resultados de IntLocProb aplicado a Teste1

Ponto de Satisfação	Desvio Padrão	Número Médio de Perguntas
0.0000	0.0000	6.49(100.00%)
0.0100	0.0289	4.08(62.82%)
0.0200	0.0361	3.94(60.71%)
0.0400	0.0480	3.61(55.65%)
0.0600	0.0534	3.54(54.56%)
0.0800	0.0472	3.44(52.93%)
0.1000	0.0570	3.21(49.40%)
0.1500	0.0804	3.11(47.86%)
0.2000	0.1036	2.87(44.28%)
0.2500	0.1429	2.57(39.55%)
0.3000	0.2515	2.06(31.76%)
0.4000	0.4246	1.37(21.17%)
0.5000	0.4743	1.11(17.12%)
0.6000	0.7369	0.13(1.93%)
0.7000	0.7393	0.13(1.93%)
0.8000	0.7378	0.13(1.93%)
0.9000	0.7240	0.00(0.00%)

FIGURA B.3 - Resultados de IntLocGC aplicado a Teste1

Ponto de Satisfação	Desvio Padrão	Número Médio de Perguntas
0.0000	0.0000	6.43(100.00%)
0.0100	0.2855	3.76(58.56%)
0.0200	0.2533	3.92(61.09%)
0.0400	0.2830	3.58(55.64%)
0.0600	0.2964	3.36(52.33%)
0.0800	0.2654	3.26(50.78%)
0.1000	0.3287	3.09(48.05%)
0.1500	0.3891	2.42(37.44%)
0.2000	0.3794	1.63(25.37%)
0.2500	0.5331	1.44(22.37%)
0.3000	0.5427	0.99(15.37%)
0.4000	0.6122	0.91(14.20%)
0.5000	0.6896	0.35(5.45%)
0.6000	0.7000	0.23(3.50%)
0.7000	0.6646	0.20(3.11%)
0.8000	0.6792	0.15(2.33%)
0.9000	0.7207	0.00(0.00%)

FIGURA B.4 - Resultados de CegLocGC aplicado a Teste1

Ponto de Satisfação	Desvio Padrão	Número Médio de Perguntas
0.0000	0.00000	11.96(100.00%)
0.0001	0.23507	9.88(82.64%)
0.0002	0.27020	9.42(78.78%)
0.0005	0.31125	7.77(64.95%)
0.0010	0.32633	6.69(55.95%)
0.0050	0.39322	3.31(27.65%)
0.0100	0.37737	2.65(22.19%)
0.0200	0.39323	2.54(21.22%)
0.0400	0.39338	1.81(15.11%)
0.0600	0.41041	1.23(10.29%)
0.0800	0.39260	1.23(10.29%)
0.1000	0.41610	0.96(8.04%)
0.1500	0.42792	0.73(6.11%)
0.2000	0.40178	0.50(4.18%)
0.2500	0.40195	0.42(3.54%)
0.3000	0.40814	0.42(3.54%)
0.4000	0.40667	0.15(1.29%)
0.5000	0.41272	0.15(1.29%)
0.6000	0.40608	0.15(1.29%)
0.7000	0.41399	0.15(1.29%)
0.8000	0.40739	0.15(1.29%)
0.9000	0.40546	0.15(1.29%)

FIGURA B.5 - Resultados de IntGlobProb aplicado a SINDROMUS

Ponto de Satisfação	Desvio Padrão	Número Médio de Perguntas
0.0000000	0.00000	11.96(100.00%)
0.0000001	0.10606	11.19(93.54%)
0.0000002	0.14259	10.95(91.55%)
0.0000005	0.16160	10.77(90.06%)
0.0000010	0.21243	10.46(87.44%)
0.0000050	0.27325	8.93(74.65%)
0.0000100	0.27659	8.04(67.19%)
0.0000200	0.29131	6.47(54.09%)
0.0000400	0.28981	5.29(44.24%)
0.0000600	0.29933	4.36(36.41%)
0.0000800	0.30457	4.14(34.61%)
0.0001000	0.31738	3.80(31.79%)
0.0002000	0.31231	3.11(26.01%)
0.0005000	0.31834	2.43(20.34%)
0.0010000	0.33210	2.08(17.42%)
0.0050000	0.33775	1.23(10.32%)
0.0100000	0.33502	1.05(8.80%)
0.0200000	0.34687	0.84(7.03%)
0.0400000	0.35658	0.49(4.11%)
0.0600000	0.37562	0.14(1.15%)
0.1000000	0.39320	0.14(1.17%)
0.2000000	0.38135	0.13(1.05%)
0.2500000	0.38553	0.11(0.96%)
0.3000000	0.39258	0.09(0.78%)
0.4000000	0.40307	0.00(0.00%)

FIGURA B.6 - Resultados de IntLocProb aplicado a SINDROMUS

Ponto de Satisfação	Desvio Padrão	Número Médio de Perguntas
0.0000	0.00000	11.74(100.00%)
0.0001	0.05516	7.28(62.00%)
0.0002	0.06200	6.53(55.59%)
0.0005	0.08144	5.46(46.51%)
0.0010	0.10078	4.87(41.47%)
0.0050	0.15777	3.56(30.34%)
0.0100	0.17556	3.35(28.53%)
0.0200	0.18602	3.22(27.42%)
0.0400	0.19939	3.05(26.02%)
0.0600	0.20344	3.01(25.65%)
0.0800	0.21260	2.83(24.13%)
0.1000	0.23734	2.65(22.57%)
0.1500	0.26063	2.46(20.96%)
0.2000	0.27076	2.32(19.73%)
0.2500	0.28682	2.21(18.86%)
0.3000	0.29543	2.08(17.68%)
0.4000	0.31885	1.72(14.67%)
0.5000	0.37564	0.20(1.70%)
0.6000	0.38288	0.07(0.57%)
0.7000	0.38331	0.03(0.26%)
0.8000	0.38472	0.02(0.20%)
0.9000	0.38018	0.02(0.13%)

FIGURA B.7 - Resultados de IntLocGC aplicado a SINDROMUS

Ponto de Satisfação	Desvio Padrão	Número Médio de Perguntas
0.0000	0.00000	11.74(100.00%)
0.0001	0.10252	7.38(62.88%)
0.0002	0.10778	6.89(58.72%)
0.0005	0.12817	6.27(53.43%)
0.0010	0.14150	5.81(49.53%)
0.0050	0.18448	4.69(39.92%)
0.0100	0.19414	3.92(33.41%)
0.0200	0.20345	3.78(32.24%)
0.0400	0.21549	3.64(31.02%)
0.0600	0.21394	3.60(30.69%)
0.0800	0.21844	3.49(29.73%)
0.1000	0.23945	3.29(28.04%)
0.1500	0.24973	3.03(25.84%)
0.2000	0.25854	2.82(24.03%)
0.2500	0.27432	2.57(21.89%)
0.3000	0.28576	2.33(19.80%)
0.4000	0.31445	1.81(15.45%)
0.5000	0.37628	0.20(1.71%)
0.6000	0.38280	0.07(0.57%)
0.7000	0.38328	0.03(0.26%)
0.8000	0.38472	0.02(0.20%)
0.9000	0.38018	0.02(0.13%)

FIGURA B.8 - Resultados de CegLocGC aplicado a SINDROMUS