Karina Dutra de Carvalho

Representação de Conhecimento Usando a Teoria de Dempster-Shafer

Monografia de Graduação apresentada ao Departamento de Ciência da Computação da Universidade Federal de Lavras como parte das exigências da disciplina Projeto Orientado para obtenção do título de Bacharel em Ciência da Computação.

Orientador
Prof. Joaquim Quinteiro Uchôa

Lavras
Minas Gerais - Brasil
2001

Karina Dutra de Carvalho

Representação de Conhecimento Usando a Teoria de Dempster-Shafer

Monografia de Graduação apresentada ao Departamento de Ciência da Computação da Universidade Federal de Lavras como parte das exigências da disciplina Projeto Orientado para obtenção do título de Bacharel em Ciência da Computação.

Aprovada em *29 de Julho de 2001*

Prof. Antônio Maria Pereira de Resende

Prof. Jones Oliveira de Albuquerque

Prof. Joaquim Quinteiro Uchôa
(Orientador)

Lavras
Minas Gerais - Brasil

À minha mãe, Maria do Carmo e
ao meu pai, Antonio Dutra,
fonte inesgotável de amor e dedicação.

Resumo

A Teoria de Dempster-Shafer (TDS), ou Teoria da Evidência, é um dos vários formalismos existentes que podem ser adotados para o tratamento de incertezas em sistemas baseados em conhecimento. Esse trabalho apresenta a TDS, seus conceitos básicos e fundamentos. São abordados com detalhes a atribuição de probabilidade básica, função de crença, função de comunalidade, plausibilidade, intervalo de crença, regra de combinação de Dempster e peso de conflito. Além disso, a TDS é avaliada como medida de incerteza, de acordo com os critérios propostos por Walley em [Walley (1996)]. É também mostrada uma comparação entre a TDS e algumas outras teorias feita por [Ng & Abramsom (1990)] quando do seus usos em sistemas especialistas.

Agradecimentos

Deus por ter me concedido a sabedoria do entendimento.

Professor Joaquim Quinteiro Uchôa, pela dedição e incentivo durante a orientação deste trabalho.

Aos meus familiares por ter contribuído de forma exemplar para minha formação.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Assim como acontece com qualquer software, Sistemas Baseados em Conhecimento (SBCs) devem ser capazes de representar, manipular e comunicar informações. É fato que tais sistemas devem estar preparados para modelar e tratar dados considerados imperfeitos. Muitas vezes o que convenciona-se chamar de dados imperfeitos abrange dados imprecisos, inconsistentes, parcialmente ignorados e mesmo incompletos. Como comentado em [Bonissone (1991), p.854]:

> a presença da incerteza em sistemas baseados em conhecimento pode se originar de várias fontes: da confiabilidade parcial que se tem na informação, da imprecisão inerente à linguagem de representação na qual a informação é expressa, da não completeza da informação e da agregação sumarização da informação que provêm de múltiplas fontes.

Existem vários modelos formais disponíveis para o tratamento de incertezas; apesar disso, muitas vezes o tratamento de incerteza em SBCs tem sido feito através de abordagens *ad hoc*, baseadas em representações e combinações de regras que não estão subsidiadas por uma teoria bem fundamentada e tampouco têm o respaldo de uma semântica bem definida.

Deve ser lembrado também que problemas relacionados com incertezas acontecem em todo sistema baseado em conhecimento. Durante o projeto de bases de conhecimento, por exemplo, deve se ter sempre em mente que o conhecimento com o qual se trabalha raramente está completo ou é exato. Como comentado em [Ng & Abramsom (1990), p.30]:

> se toda informação pudesse ser representada de maneira completa e

precisa, qualquer sistema robusto de inferência lógica poderia ser utilizado para a extração de conclusões válidas.

É preciso que se implemente maneiras de lidar com essa situação. Assim, bases de conhecimento se constituem numa das principais fontes de informações incertas em sistemas baseados em conhecimento. Entre as abordagens mais tradicionais existentes para a modelagem e tratamento de incertezas, encontram-se:

- fatores de certeza [Shortliffe & Buchanan (1975)];
- teoria de Dempster-Shafer [Dempster (1967)] [Shafer (1976)];
- regra de Bayes modificada [Duda et al. (1976)];
- teoria da possibilidade [Zadeh (1978)];
- raciocínio default [Reiter (1980)];
- regra de Bayes [Pearl (1982)];
- teoria de conjuntos aproximados [Pawlak (1982)][Uchôa (1998)];
- teoria do endorsements [Cohen (1985)];

Este trabalho tem por objetivo principal apresentar e discutir a TDS, bem como verificar sua usabilidade para representação de incerteza em SBCs. Ele encontra-se dividido da seguinte forma: no Capítulo 2 é abordado tratamento de incerteza em SBCs. Sendo apresentado a Teoria da Probabilidade, Teoria da Possibilidade e Fatores de Certeza nas Seções 2.1, 2.2 e 2.3, respectivamente. O Capítulo 3 apresenta os principais conceitos da Teoria de Dempster-Shafer, bem como comenta sobre funções de crença Bayeseanas, mostrando como alguns conceitos da Teoria de Bayes podem ser expressos pela TDS. Este c'ápitulo apresenta alguns teoremas que podem ser encontrados no livro do [Shafer (1976)]. No Capítulo 4 é realizada uma avaliação do uso da Teoria de Dempster-Shafer para representação de incerteza em SBCs. Por último, o Capítulo 5 apresenta as principais conclusões sobre a utilização da TDS.

Além disso, o Apêndice A apresenta trechos principais de uma implementação em linguagem C++ dos elementos básicos da Teoria de Dempster-Shafer facilitando a compreensão dessa teoria.

Capítulo 2

Tratamento de Incerteza em Sistemas Baseados em Conhecimento

2.1 Sistemas Baseados em Conhecimento

SBCs são definidos formalmente como programas de computador que resolvem problemas utilizando conhecimento representado explicitamente e que, não fosse essa representação, exigiriam um especialista humano no domínio do problema para a sua solução. Conhecimento e processo de resolução de problemas são pontos críticos e essenciais durante o desenvolvimento de um SBC. A arquitetura básica de um SBC pode ser visualiza da na Figura 2.1.

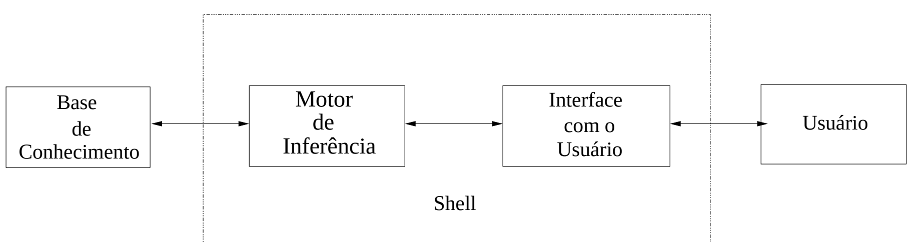

Figura 2.1: Arquitetura básica de um SBC

Um SBC possui, então, três módulos principais, a saber:

- *Base de Conhecimento* (BC): contém o conhecimento específico do domínio da aplicação. É composto de fatos sobre o domínio, regras que descrevem relações no domínio e métodos e heurísticas para resolução de problemas no domínio.

- *Motor de Inferência* (MI): mecanismo responsável pelo processamento do conhecimento da BC, utilizando-se de alguma linha de raciocínio. Implementa as estratégias de inferência e controle do SBC. Quando o conhecimento do SBC está expresso como regras, as estratégias de controle empregadas pelo MI normalmente são encadeamento para trás (*backward chaining*) ou encadeamento para frente (*forward chaining*). Quando o MI usa a estratégia de encadeamento para trás, ele pesquisa uma lista de hipóteses e procura reunir evidências para viabilizar a conclusão da verdade de alguma(s) dela(s). Esta estratégia corresponde à pergunta: *É possível provar as hipóteses a partir dos dados disponíveis?* Se a estratégia de controle for encadeamento para frente, o MI parte dos dados e, com base nas regras de conhecimento, deduz outras asserções procurando chegar à solução do problema. Essa estratégia corresponde à pergunta: *O que é possível concluir a partir dos dados disponíveis?*

- *Interface com o usuário* (IU): módulo responsável pela comunicação entre o usuário e o sistema. Deve fornecer, também, justificativas e explicações referentes às conclusões obtidas na BC, bem como do raciocínio utilizado.

É fato que os aspectos principais de um SBC são a capacidade de resolver problemas que demandam informação referente a um determinado domínio de conhecimento e a capacidade de suas decisões ao usuário, ou seja, *porque e como* se chegou a uma determinada solução. De uma maneira geral as perguntas *por que* refere-se a qual conhecimento respalda a conclusão; a pergunta *como*, por sua vez, referem-se aos passos de raciocínio seguidos para determinar a solução do problema. Esta característica é especialmente necessária quando o SBC lida com domínios incertos; a explição pode, de certa forma aumentar o grau de confiança que o usuário deposita no sistema, ou então, ajudá-lo a encontrar alguma falha no raciocínio do sistema.

Uma outra característica freqüentemente necessária é a habilidade de lidar com incertezas e informações incompletas: a informação a respeito do problema a ser resolvido pode estar incompleta ou ser parcialmente confiável, bem como as relações no domínio do problema podem ser aproximadas. Espera-se, também, que um SBC seja flexível o suficiente para permitir facilmente, a acomodação de novo conhecimento.

Uma importante subclasse dos SBCs é a constituída pelos sistemas especialistas. Um sistema especialista pode ser definido como um SBC que resolve problemas específicos do mundo real, problemas esses que requerem considerável habilidade, conhecimento e heurísticas para sua resolução. Na década passada já tínhamos sistemas especialistas totalmente desenvolvidos que foram projetados para darem uma importante contribuição na área da medicina, da engenharia, comercial e financeira [Ng & Abramsom (1990)]. Entretanto, projetos mais complexos de sistemas especialistas ainda são problemáticos, o que gera a seguinte conclusão - gerenciamento de incerteza é ainda um caso em discussão.

Um sistema especialista é um programa de computador que possue três características importantes [Ng & Abramsom (1990)]:

1. O banco de dados é projetado com a ajuda de um especialista humano;

2. O domínio do problema é bastante restrito;

3. É esperado que ele tenha uma performace equivalente da dos especialistas humanos.

Estes sistemas lidam constantemente com incerteza porque todo domínio completo e bom de sistemas contém dados que são inconsistentes, imprecisos e até mesmo incompletos. Para tratar estes dados são usados os modelos de tratamento de incerteza.

Um modelo para tratamento de incertezas fornece maneiras de expressar incertezas associadas a hipóteses, bem como maneiras de combinar várias incertezas em uma única. Na literatura podem ser encontradas diferentes abordagens que viabilizam o raciocínio incerto, geralmente como alternativas ao raciocínio probabilístico. Como comentado em [Heckerman (1986)], muitos modelos são justificados devido a limitações práticas e teóricas da teoria da probabilidade. Já outros pesquisadores argumentam afirmando que métodos probabilísticos são suficientes para o tratamento de incertezas em todos os casos. O certo é que o tratamento de incertezas através de uma abordagem baseada em probabilidades exige que uma quantidade razoável de dados esteja disponível. Devido a este fato, algumas aproximações/modelos relativamente menos poderosos para o tratamento de incertezas foram propostos com o objetivo de reduzir tal exigência [Santos & Nicoletti].

Com o objetivo de apresentar alguns mecanismos de tratamento de incerteza em SBCs, esse capítulo fará uma rápida abordagem em probabilidade subjetiva, a teoria da possibilidade e modelo de fatores de certeza, que também são modelos de tratamento de certeza. A teoria de Dempster-Shafer, objeto principal desse trabalho, será abordado no próximo capítulo.

2.2 Teoria da Probabilidade Subjetiva

O conceito de probabilidade surgiu há séculos com a necessidade de introduzir as palavras “talvez”, “provável”, “chance” e “acaso” a nosso cotidiano. Entretanto, a teoria matemática da probabilidade foi formulada por volta do ano de 1660, quando foi formalizada a sua terminologia. Então a teoria da probabilidade é considerada uma das teorias matemáticas mais antigas que tratam de incerteza.

SBCs têm utilizado o raciocínio probabilístico de Bayes, derivado da fórmula de probabilidade condicional, uma vez que tal modelo permite coletar evidências durante o uso do sistema e redirecionar o processo de inferência à medida que novas evidências surgem. O desempenho de um sistema que utiliza esta teoria para modelar a incerteza dependerá, na maioria das vezes, do domínio do problema o qual o sistema estiver modelado [Santos & Nicoletti].

Existem várias interpretações para o termo *probabilidade*, existindo três visões dominantes:

- *Objetiva:* probabilidade que mede a relação de ocorrências de observações para uma proposição. Em outras palavras, a lei garante que com grandes números de dados observados, a porcentagem de um evento ocorrer aproxima-se da probabilidade objetiva.

- *Subjetiva:* probabilidade que mede a confiança que um indivíduo tem de que uma proposição particular seja verdade. Existe a possibilidade de que dois indivíduos acreditem que uma mesma evidência seja verdade mas com diferentes graus de confiança. O termo “Bayesiano” é freqüentemente usado como sinônimo de probabilidade subjetiva.

- *Lógica:* probabilidade que mede o grau que uma pessoa fixou para um conjunto de proposições. Proponentes geralmente consideram esta visão como extensão de lógica.

Agora será apresentado o Teorema de Bayes, uma vez que essa equação é base para a inferência probabilística utilizada em vários SBCs:

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}.$$

À primeira vista, a regra de Bayes não parece ser de grande utilidade, uma vez que para aplicá-la no cálculo de uma probabilidade condicional $p(A|B)$, são necessárias uma probabilidade condicional - $p(B|A)$ - e duas probabilidades incondicionais - $p(A)$ e $p(B)$. Entretanto, tal fórmula é útil na prática porque, em muitas situações, existem boas estimativas para $p(B|A)$, $p(A)$, $p(B)$ e existe a necessidade de se obter $p(A|B)$. Em diagnósticos médicos, por exemplo, freqüentemente as probabilidades condicionais de relações causais são conhecidas, e se deseja derivar um diagnóstico.

Existe um processo de normalização da regra de Bayes, denominado *normalização*, o que permite que sejam levados em conta todos os outros possíveis diagnósticos com relação a determinado sintoma, sem que seja necessário computar as razões relativas entre cada par de diagnósticos e sem que seja preciso o conhecimento das probabilidades a priori dos sintomas.

Considere uma situação em que todas as regras de um SBC são expressas na forma: ”SE <H é verdade> ENTAO <E será observada com probabilidade p>”, onde H denota uma hipótese e E uma evidência.

Se H é observada, essa regra afirma que a probabilidade de que o evento E ocorreu é p. Entretanto o que acontece se o valor de H é desconhecido e a evidência E for observada?

A equação a seguir estabelece como calcular a probabilidade de H ser verdade dado que E é observada:

$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E|H) \times p(H) + p(E|\neg H) \times p(\neg H)}.$$

Em SBCs, as probabilidades necessárias à solução do problema são fornecidas por especialistas humanos e armazenados na base de conhecimento. Essas probabilidades são as probabilidades *a priori* para todas possíveis hipóteses H $(p(H))$ bem como as probabilidades condicionais de se observar uma evidência E dada uma hipótese H $(p(E|H))$, para todas H e E. Elas são necessárias para o cálculo da probabilidade *a posteriori* da hipótese A, dada a observação da evidência E; $p(H|E)$ indica o quanto se pode acreditar na hipótese H, dado que a evidência E foi observada.

Como discutido em [Ng & Abramsom (1990)], no desenvolvimento de SBCs para domínios médicos, um médico especialista deve fornecer as probabilidades a priori de todas as possíveis doenças relativas ao domínio médico em questão. Além disso, a probabilidade condicional para a observação de um sintoma, dada uma doença específica, deve também ser fornecida para todos os pares de sintoma e doença (assumindo que todos os sintomas são condicionalmente independentes dada uma doença). Se os sintomas não forem condicionalmente independentes, as probabilidades condicionais para a observação das várias possíveis combinações de sintomas, dada uma doença específica, devem também ser fornecidas, para cada possível doença. A suposição que os sintomas sejam condicionalmente independentes é fundamental para tornar viável a utilização da regra de Bayes.

Em um domínio de diagnóstico médico, por exemplo, um SBC que viabiliza o diagnóstico de 100 doenças, considerando 500 possíveis sintomas (evidências), deve dispor de pelo menos 50.100 valores probabilísticos (50.000 probabilidades condicionais e 100 probabilidades *a priori*). Além disso, as suposições fundamentais para a aplicação da regra de Bayes devem ser garantidas: todas as doenças são mutuamente exclusivas, todos os sintomas são condicionalmente independentes dada uma doença, e todos sintomas estão restritos a dois valores (verdadeiro e falso). Se forem incorporados dependência dos sintomas e possibilidade de múltiplas doenças (não mutuamente exclusivas), o número de valores probabilísticos aumentaria consideravelmente.

Infelizmente, em problemas do mundo real, a suposição de independência condicional raramente é válida [Szolovits (1978)] e a suposição de exclusividade mútua e a exaustividade de categorias de doenças é usualmente falsa. Além desses, para o uso do modelo de Bayes, existe o problema da avaliação de um número extraordinariamente grande de valores probabilísticos.

2.3 Teoria da Possibilidade

A teoria da possibilidade [Zadeh (1978)] foi desenvolvida por Lotfi Zadeh como uma extensão da teoria dos conjuntos *fuzzy* [Zadeh (1965)]. A dificuldade em representar informações incompletas e vagas usando teoria da probabilidade motivaram as pesquisas desta teoria.

Em um sistema especialista, a base de conhecimento é formada em grande parte por conhecimentos humanos, os quais são imprecisos e qualitativos. Freqüentemente os limites entre hipóteses competitivas são vagamente definidas. Para expressar conhecimento sobre tais problemas, especialistas humanos usam termos tais como ”provavelmente” para descreverem ocorrências de evidências e hipóteses. Quando este tipo de perícia é passada para probabilidade, a imprecisão é usualmente perdida, e a idéia é representar com pontos de valores específicos. Zadeh desenvolveu a teoria da possibilidade para expressar esses termos vagos com precisão e exatidão.

A teoria da possibilidade substitui a lógica binária de probabilidade por lógica de multivalores. Em teoria da probabilidade existem apenas dois graus de pertinência para os elementos de um conjunto, a pertinência zero e a pertinência um. Neste caso um evento ocorre ou ele não ocorre. Na teoria da possibilidade (asssim como lógica fuzzy, Dempster-Shafer e teorias correlatas) a pertinência pertence ao espectro contínuo ($\mathbb{R}$) dentro do intervalo $[0, 1]$. Neste caso um elemento (ou evento) *e* pode ocorrer, não ocorrer ou ainda outras variações do tipo:

- quase ocorrer;
- quase não ocorrer;
- ocorrer mais ou menos;
- ter boa ocorrência;
- ocorrência ruim.

Para estas variações verifica-se que $\{p(e) \in \mathbb{R} \ / \ p(e) > 0 \ \ e \ \ p(e) < 1\}$.

Como esta teoria é baseada na teoria de conjuntos fuzzy, se faz necessário uma rápida visão das terminologias, propriedades e definições dos conjuntos *fuzzy*. Dado U um conjunto de objetos. Um conjunto *fuzzy* é uma classe de objetos com grau contínuo de associação em U. Suponha que A é um subconjunto fuzzy de U caracterizado por uma função de pertinência $\mu_a(u)$ a qual associa um número real entre $[0, 1]$ com cada elemento de $u \in U$. O valor de $\mu_a(u)$ representa o grau de pertinência de u em A.

Assim, a diferença entre conjuntos *fuzzy* e um conjunto normal é os possíveis valores de μ_a. Em um conjunto normal, a função de pertinência pode somente assumir os valores 0 e 1: $\mu_a(u) = 1$ denota a pertinência e $\mu_a(u) = 0$ denota não pertinência. Estas duas teorias também diferem no número fundamental de tarefas, sem relação direta entre elas - necessidade de possibilidade alta não implica probabilidade alta, e vice-versa.

Um conjunto fuzzy é dito ser vazio se e somente se a função de relação é zero para todos elementos de U. Dois conjuntos fuzzy A e B são iguais se e somente se $\mu_a(u) = \mu_b(u)$ para todo u em U. O complemento de um conjunto fuzzy é denotado por $\neg A$, e a função de associação é definida por $\mu_{\neg A} = 1 - \mu_A$. Os conjuntos fuzzy ainda têm as propriedades de união e de intercecção. A união de dois conjuntos A e B é o conjunto fuzzy C, com $\mu_C(u) = max[\mu_A(u), \mu_B(u)]$ para todo $u \in U$. Similarmente, a intercecção de dois conjuntos é um outro conjunto fuzzy C, onde $\mu_C(u) = min[\mu_A(u), \mu_B(u)]$ para todo $u \in U$. As operações $\cap$

e $\cup$ são ambas comutativas e associativas. A lei de De Morgan e distributividade também são válidas para conjuntos fuzzy.

A questão que se coloca agora é como fazer propagação de crença. O paradigma que sugere essa propagação é conhecido como teoria da possibilidade, a qual trabalha principalmente com distribuições diretamente relacionadas com a função de pertinência. Se A é um subconjunto fuzzy de U caracterizado pela função de pertinência $\mu_A : U \rightarrow [0,1]$, a proposição "$X$é$A$" associa a distribuição de possibilidade Π_X tal que $\Pi_X = A$. Correspodentemente, a função de distribuição de possibilidade associada com X, μ_X, é igual a função de pertinência de A, isto é, $\pi_X = \mu_A$.

Com base nessas proposições, define-se a medida de possibilidade $Poss\{x \in A\}$, que é a possibilidade que o valor de x pertença a A, e é expressa como $Poss\{x \in A\} = max_{u \in A}[\pi_x(u)]$.

2.4 Fator de Certeza

O Modelo de fatores de certeza é um modelo "quase-probabilístico" para tratamento de incertezas, que foi especialmente desenvolvido para o sistema MYCIN, por Shorliffe e Buchanan [Shortliffe & Buchanan (1975)]. Esse modelo, desde sua criação, tem sido adotado por vários sistemas especialistas baseado em regras, construídos após o MYCIN. Mesmo não sendo bem fundamentado matematicamente, na prática oferece resultados satisfatórios. Alguns autores justificam que o sucesso relativo do modelo se deve à sua simplicidade computacional.

Em um modelo de tratamento de incertezas, em um sistema de regras de produção, usando fatores de certeza, o especialista humano deve expressar sua incerteza com relação à hipótese H de uma regra de produção

"Se E(EVIDÊNCIA) então H(HIPÓTESE)",
com um fator de certeza $CF[H, E]$.

Fatores de certeza representam a confiança que se tem em uma evidência. Existem diferentes maneiras de se representar fatores de certeza e deve ficar claro que fatores de certeza não são probabilidades, são medidas informais de confiança ou certeza com relação à uma evidência. Eles representam o grau de crença de que a evidência é, de fato, verdade.

Um fator de certeza positivo é associado à hipótese H dada a evidência E se existir algum grau de confirmação em H, dada a observação E. O fator de certeza $CF[H, E] = 1$ indica que a ocorrência de E prova completamente a hipótese H.

O fator de certeza negativo é adotado pelo especialista se a observação da evidência E desconfirma a hipótese H. Um fator de certeza zero é atribuído se a observação de E não influência a confiança na hipótese H.

Como comentado anteriormente, em um SBC que se faz uso de fatores de certeza, a base de conhecimento consiste de um conjunto de regras na forma:

Se <EVIDÊNCIA> então <HIPÓTESE> (CF),

onde CF é um valor que representa a mudança da crença na HIPÓTESE, dada a observação de EVIDÊNCIA. O fator de certeza é, na realidade, uma atualização da crença.

Um aumento da crença na hipótese é indicado por um fator de certeza positivo ou uma diminuição, por um fator de certeza negativo. A definição do fator de certeza de uma hipótese dada a evidência E envolve o cálculo de duas funções intermediárias:

- MB[H,E]: mede o grau de aumento da crença na hipótese H, se E for observada.
- MD[H,E]: mede o grau de aumento da descrença na hipótese H, se E for observada.

Ambas as medidas, MB e MD, são definidas em função de:

- p(H): probabilidade *a priori* de H;
- p(H|E): probabilidade *a posteriori* de H ou probabilidade condicional, dada a evidência E.

As noções de crença e descrença foram criadas com o objetivo de refletir os conceitos intuitivos de confirmação e desconfirmação. A medida do aumento de crença MB é uma função que associa dois argumentos: a hipótese H e uma evidência E a um valor no intervalo $[0, 1]$. A MB de uma hipótese H dada a evidência E pode ser pensada como a diminuição proporcional da descrença em H dada E e é equacionada em função de probabilidades como:

$$MB[H, E] = \begin{cases} (1), & \text{se } p(H) = 1 \\ \frac{max[p(H|E), p(H)] - p(H)}{max[0,1] - p(H)}, & \text{caso contrário.} \end{cases}$$

De maneira análoga, a medida do aumento de descrença MD é a diminuição proporcional da crença em H, como resultado da evidência E. A medida do aumento de descrença também é uma função que associa dois argumentos: a hipótese H e uma evidência E a um valor no intervalo $[0, 1]$ e é equacionada em função de probabilidades como:

$$MD[H, E] = \begin{cases} (1), & \text{se } p(H) = 1 \\ \frac{min[p(H|E), p(H)] - p(H)}{min[0,1] - p(H)}, & \text{caso contrário.} \end{cases}$$

O valor de MB e MD é um valor entre zero e um. Assim sendo, apenas um dos dois valores pode ser não zero para uma determinada hipótese H e evidência E e são deduzidos a partir da medida $CF[H|E]$. As medidas de crença e descrença capturam explicitamente a noção de que uma evidência não pode favorecer e desfavorecer uma única hipótese, ou seja, se $MD[H|E] > 0$ então $MB[H|E] = 0$ e vice-versa. O fator de certeza $CF[H|E]$ é definido por:

$$CF = \frac{MB - MD}{1 - min[MB, MD]}.$$

O valor de CF é um número real variando entre -1 e +1. O valor -1 indica a confirmação da negação de H e o valor +1 indica a confirmação de H. Eles são fornecidos pelo especialista humano quando a regra é adicionada ao sistema.

A interpretação original do CF como um incremento (ou decremento) probabilístico, não pode ser preservada, após os CF's terem sido agregados usando as funções heurísticas de combinação fornecidas pelo MYCIN. Essas funções de combinação são aproximações do procedimento *bayesiano* clássico de atualização, onde um termo é ignorado. Além disso, a suposição de mútua independência de evidências é necessária para o uso correto da abordagem.

2.5 Comentários Finais

Este capítulo apresentou alguns dos modelos mais utilizados na literatura para representação e tratamento de conhecimento incerto. O objetivo dessa apresentação é permitir uma comparação dessas teorias com a TDS que será apresentada no Capítulo 4.

Capítulo 3

Elementos da Teoria de Dempster-Shafer

A Teoria de Dempster-Shafer se originou com o trabalho de Arthur Dempster sobre probabilidades inferiores e superiores [Dempster (1967)], [Dempster (1967a)] e teve continuidade com os trabalhos de Glen Shafer [Shafer (1976)] que refinou e estendeu as idéias de Dempster.

O investimento no modelo Dempster-Shafer para o tratamento de incertezas em sistemas baseados em conhecimento foi motivado principalmente por problemas encontrados na modelagem da incerteza usando métodos puramente probabilísticos e também pela falta de embasamento matemático do modelo Fator de Certeza do MYCIN. Como comentado em [Gordon & Shortliffe (1984), P.272]:

> a vantagem da Teoria de Dempster-Shafer sobre as abordagens anteriores está na habilidade deste método em modelar o afunilamento do conjunto de hipóteses, à medida em que se acumulam evidências: este procedimento reflete o processo que caracteriza o raciocínio em diagnósticos e o raciocínio especializado em geral.

A Teoria de Dempster-Shafer tem muitos aspectos em comum com o modelo de Fator de Certeza, mas mesmo assim, ao contrário deste modelo, a TDS é bem fundamentada matematicamente. A regra de combinação do modelo de Fator de Certeza é, na realidade, especializações da regra de combinação da TDS.

A TDS permite que funções de crença possam ser combinadas, produzindo assim novas funções de crença num procedimento que independe da ordem na qual as evidências surgem, mas que, entretanto, exige que as *hipóteses primitivas* consideradas sejam mutuamente exclusivas e exaustivas. A partir destas hipóteses primitivas, também denominadas de *singletons*, é possível construir hipóteses mais elaboradas que não são mutuamente exclusivas ou exaustivas.

A partir de uma única coleção de evidências, usando a TDS, vários conjuntos alternativos de hipóteses podem ser derivados. A cada um desses conjuntos está associado um intervalo de confiança. Quando existe a validade de uma determinada hipótese, a TDS permite que possam ser consideradas todas as informações disponíveis.

Como a TDS atribui valores de crença a subconjuntos e a cada elemento do conjunto de hipóteses, essa teoria tem condições de refletir mais precisamente o processo de acúmulo de evidências.

3.1 Conceitos Básicos da TDS

Neste capítulo serão definidos três conceitos básicos necessários para o entendimento do capítulo posterior.

Domínio do problema, também chamado de *frame de discernimento*, ou ainda *quadro de discernimento* é o conjunto das hipóteses primitivas e é notado por Θ.

EXEMPLO 1 *[Uchôa & Nicoletti (1997)] Suponha que um paciente apresente manchas vermelhas pelo corpo e que isso seja sintoma de qualquer dos seguintes problemas: alergia $\{a\}$, intoxicação $\{i\}$, sarampo $\{s\}$, rubéola $\{r\}$.*

O domínio do problema é $\Theta = \{a, i, s, r\}$.

A TDS assume para qualquer domínio do problema Θ que:

1. Θ é exaustivo, no sentido de ser completo, ou seja contêm todas possíveis hipóteses primitivas;

2. as hipóteses primitivas em Θ são mutuamente exclusivas.

Suponha que agora um outro sintoma, ou seja evidência, considerada pelo médico aponte que:

- para um diagnóstico de *reação orgânica*, definida no exemplo como o conjunto $\{a, i\}$, ou então

- para um diagnóstico de *infecção*, definida no exemplo como o conjunto $\{s, r\}$.

Se o médico observar uma evidência que confirma com um determinado grau, por exemplo, o diagnóstico *reação orgânica*, será atribuído ao conjunto $\{a, i\}$ uma quantidade de crença proporcional ao grau observado de confirmação de evidência.

Pode haver também uma nova evidência que, por exemplo, exclua alergia do diagnóstico, o que pode ser tratada como uma evidência que confirme o resto do conjunto de hipóteses $\{i, s, r\}$. Este novo subconjunto de hipóteses de Θ pode ser visto como um nova hipótese, formada pela disjunção de seus elementos, uma vez que como citado anteriormente, hipóteses primitivas são mutuamente exclusivas.

O papel do domínio do problema Θ na TDS se assemelha ao do espaço amostral Ω na Teoria de Probabilidade, entretanto a diferença é que na TDS o número de possíveis hipóteses é $2^{|\theta|}$, enquanto que na Teoria de Probabilidade é de $|\Omega|$.

A Figura 3.1 mostra as possíveis hipóteses dispostas (que compõe um reticulado, induzindo pela relação de ordem parcial $\subseteq$ existente entre elas):

Θ

{a,i,s} {a,i,r} {a,s,r} {i,s,r}

{a,i} {a,s} {a,r} {i,s} {i,r} {s,r}

{a} {i} {s} {r}

ϕ

Figura 3.1: Conjunto de todas as possíveis hipóteses obtidas de $\Theta = \{a, i, s, r\}$

Na TDS, um valor no intervalo $[0, 1]$ é associado à crença de uma hipótese. A relevância de cada evidência para cada um dos elementos de todos possíveis subconjuntos de Θ é representada por uma função chamada de *atribuição de probabilidade básica* (bpa) ou *função de massa*.

A bpa é uma generalização da função de densidade da probabilidade porque a bpa associa um número do intervalo $[0, 1]$ a todo subconjunto Θ, garantindo, entretanto, que a soma dessas atribuições seja um. Ainda, a bpa notada por *m*, por definição atribui zero ao conjunto vazio, uma vez que o conjunto vazio corresponde à hipótese falsa porque Θ é exaustivo.

Essa função representa a quantidade total de crença na evidência que aponta exatamente para um determinado conjunto de hipóteses. Assim sendo, *m* permite a atribuição de uma quantidade de crença a cada elemento do reticulado visto na Figura 3.1, e não apenas aos conjuntos $\{a\}, \{i\}, \{s\} \quad e \quad \{r\}$ como acontece na função de densidade da probabilidade.

Sendo a crença total um e A qualquer elemento de 2^{θ} temos por definição que a quantidade de $m(A)$ é a medida daquela parte da crença total, que é atribuída exclusivamente a A e posteriormente não pode ser subdividida entre os subconjuntos de A e não inclui a parte da crença atribuída a subconjuntos de A [Gordon & Shortliffe (1984)]. Formalmente, se Θ é um *domínio de problema*, então $m : 2^{\Theta} \rightarrow [0, 1]$ é chamada de *atribuição básica de probabilidade* se satisfaz:

1. $m(\varnothing) = 0$;
2. $m(A) \geq 0, \forall A \in 2^{\Theta}$;
3. $\sum_{A \in 2^{\theta}} m(A) = 1$.

Após as probabilidades básicas terem sido atribuídas aos subconjuntos próprios de Θ, as que não foram atribuídas, por convenção, são chamadas de *crença não atribuída*, notada por $m(\Theta)$. Neste ponto o modelo de Bayes difere da TDS o qual atribui o resto da crença a Θ e não à negação da hipótese. Por exemplo se $m(A) = x$ e m não atribui crença a qualquer outro subconjunto de Θ, então $m(\Theta) = 1 - x$.

A TDS vê a observação de evidência contra uma hipótese apenas como evidência que suporta a negação desta hipótese, como por exemplo, se têm uma evidência que desconfirme a hipótese $\{a\}$ é equivalente a evidência que confirma a hipótese $\{i, s, r\}$, ou seja qualquer hipótese que não tenha alergia.

EXEMPLO 2 *Para o exemplo anterior, uma possível atribuição de probabilidade básica poderia ser:*

$$m(\{s\}) = 0.2;$$
$$m(\{a\}) = 0.3;$$
$$m(\{r\}) = 0.1;$$
$$m(\{a, s\}) = 0.4;$$
$$m(A) = 0, \quad para \forall \quad A \in 2^{\theta},\ A \neq \{a\},\ A \neq \{s\},\ A \neq \{r\},\ A \neq \{a, s\}.$$

EXEMPLO 3 *[Strat (1994)] Um jogo familiar de quase todos é o da roleta, como mostrado na Figura 3.2. O jogo funciona da seguinte forma: A roleta é dividida em dez partes iguais, sendo que cada parte é entiquetada com uma quantia em dólares, que pode ser um, cinco, dez ou vinte dólar. Para jogar é preciso pagar uma taxa de seis dólares. Depois é só girar a roleta e receber a quantia mostrada pela seta que fica no topo da roleta. A pergunta é: quais as chances de ganhar dinheiro?*

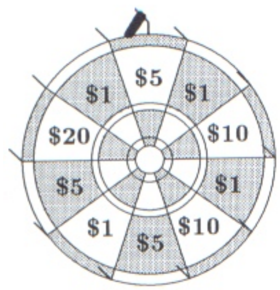

Figura 3.2: Jogo da Roleta

A análise para este problema é feita como uma representação probabilística. Então se constrói a seguinte distribuição probabilística: p($1) = 0.4, p($5) = 0.3, p($10) = 0.2 e p($20) = 0.1 O valor esperado denotado por $E(x)$ é calculado com a fórmula

$$E(x) = \sum_{x \in \Theta} x.p(x); \tag{3.1}$$

onde Θ é formado pelo conjunto dos possíveis resultados. O valor esperado da roleta é $\$5.90$ como mostra a Tabela 3.1.

Tabela 3.1: Valor esperado no jogo da roleta

x	$p(x)$	$x.p(x)$
1	0.4	0.4
5	0.3	1.5
10	0.2	2.0
20	0.1	2.0
	$E(x) =$	5.90

Como visto na Tabela 3.1, o valor esperado que é $\$5.90$ é menor que o custo para jogar a partida ($\$6.0$), o que mostra que não é viável jogar.

Agora este mesmo problema sofrerá uma leve modificação mostrada no Exemplo 4.

EXEMPLO 4 *[Strat (1994)] Suponhamos agora que a mesma roleta esteja com um de seus setores tampado, como mostra a Figura 3.3. A pergunta é quanto nós podemos estar dispostos a pagar para jogar uma partida?*

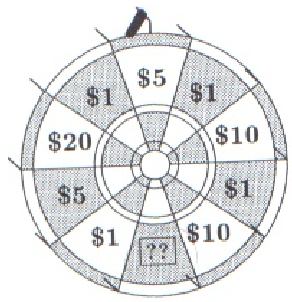

Figura 3.3: Jogo da Roleta(alterado)

Este problema é ideal para uma análise usando função de crença. O domínio desse problema Θ é $\{\$1, \$5, \$10, \$20\}$. A atribuição de probabilidade básica é a seguinte:

$$m(\{\$1\}) = 0.4;$$

$$m(\{\$5\}) = 0.2;$$

$$m(\{\$10\}) = 0.2;$$

$$m(\{\$20\}) = 0.1;$$

$$m(\{\$1, \$5, \$10, \$20\}) = 0.1.$$

Para calcular o valor esperado, utiliza-se da fórmula que se segue:

$$E(x) = [E_*(x), E^*(x)], \qquad (3.2)$$

onde

$$E_*(x) = \sum_{A_i \subseteq \Theta} inf(A_i).m_\Theta(A_i),$$

$$E^*(x) = \sum_{A_i \subseteq \Theta} sup(A_i).m_\Theta(A_i).$$

Então o intervalo do valor esperado utilizando a Equação 3.2 é:

$$E(x) = [0.4(1) + 0.2(5) + 0.2(10) + 0.1(20) + 0.1(1), 0.4(1) + 0.2(5)$$

$$+0.2(10) + 0.1(20) + 0.1(20)] = [5.50, 7.40]$$

Se o valor esperado fosse calculado utilizando a Equação 3.1 o resultado seria $\$6.30$ *como mostrado na Tabela 3.2.*

Tabela 3.2: Valor esperado no jogo da roleta utilizando a equação 3.1

x	$p(x)$	$x.p(x)$
1	0.425	0.425
5	0.225	1.125
10	0.225	2.250
20	0.125	2.500
	$E(x) =$	6.30

EXEMPLO 5 *[Stein (1993)] Considere uma situação de previsão no mercado de ações onde* $\Theta = \{NMG, -5\%, -1\%, 0\%, 1\%, 5\%, PMG\}$*, onde cada elemento é uma hipótese indicando uma mudança no preço de ações nas próximas 24 horas e os termos NMG e PMG indicam mudanças negativas e positivas muito grande, respectivamente. Considere, agora o conjunto* $H_1 = \{1\%, 5\%, PMG\}$ *que contém as hipóteses que refletem um movimento de valorização no mercado financeiro. Suponha que uma determinada regra de análise financeira suporte em* 60% *o conjunto de hipótteses* H_1*, dada uma determinada evidência. A função m então é calculada como:*

$$H_1 = \{1\%, 5\%, PMG\} \qquad m(H_1) = 0.6$$

$$\Theta \qquad m(\Theta) = 1 - 0.6 = 0.4$$

Note que Θ *contém o conjunto* H_1*, assim como o seu complemento* $\overline{H_1} = \{NMG, -5\%, -1\%, 0\%\}$*. Usando a TDS, é errado atribuir o valor 0.4 apenas a* $\overline{H_1}$*, uma vez que não existe evidência que o restante 0.4 de probabilidade de fato contradiz* H_1*. Sabe-se apenas que a evidência existente suporta* H_1 *com uma confiança de 0.6. Essencialmente, o que se está dizendo com isso é que se está*

60% confiante que a evidência observada indica uma valorização expressa em H_1. Por outro lado, sabe-se com 40% de confiança que a evidência observada, não diz nada. Atribui-se então, os 40% restante de probabilidade ao domínio do probelma, o qual contém ambos: H_1 e $\overline{H_1}$. Mais tarde, com o aparecimento de novas evidências, esses 40% podem ser reduzidos ainda mais.

Os Exemplos 6, 7 e 8 são modificações do Exemplo 1 retirados de [Uchôa & Nicoletti (1997)].

EXEMPLO 6 *Suponha que não exista evidência com relação a qualquer diagnóstico em um paciente com manchas vermelhas. A função de atribuição de probabilidade atribui 1 a $\Theta = \{a, i, s, r\}$ e 0 a qualquer outro subconjunto de Θ. O modelo de Bayes tenta representar a ignorância através de uma função que atribui 0.25 a cada hipótese primitiva, assumindo nenhuma informação a priori. É importante notar que tal atribuição implica mais informação do que realmente existe.*

EXEMPLO 7 *Suponha agora uma evidência que desconfirma o diagnóstico de sarampo, com um grau 0.7. Isso é equivalente a confirmar o não sarampo com grau de 0.7. Assim $m(\{a, r, i\}) = 0.7$, $m(\Theta) = 0.3$ e o valor de m para qualquer outro subconjunto de Θ é zero.*

EXEMPLO 8 *Suponha uma evidência que confirma o diagnóstico de sarampo, com um grau 0.6. Então $m(\{s\}) = 0.6$, $m(\Theta) = 0.4$ e m é zero em qualquer outro conjunto.*

Como observado, a TDS permite que o *peso* de várias "pequenas evidências" possa ser relevante, mesmo que nenhuma delas, sozinha, seja relevante.

3.2 Função de Crença

A *função de crença* é denotada por *bel* e atribui a todo subconjunto A de Θ, a soma das probabilidades básicas representadas pela função de atribuição de probabilidade básica m.

A quantia $m(A)$ mede a crença que se atribui exatamente a A e não o total de crença que se atribui a A. Para se obter a medida do total de crença atribuído a A, deve-se adicionar à $m(A)$ os valores $m(B)$, para todo subconjunto próprio B de

A:

$$bel(A) = \sum_{X \subseteq A} m(X). \tag{3.3}$$

Uma função $bel\colon 2^{\theta} \rightarrow [0,1]$ é chamada de *função de crença* sobre Θ se ela for dada por 3.3, relativa a alguma atribuição de probabilidade básica $m\colon 2^{\theta} \rightarrow [0,1]$. Com certeza a função de crença com estrutura mais simples é aquela obtida fazendo $m(\Theta) = 1$ e $m(A) = 0$ para todo $A \neq \Theta$. Ela tem $bel(\Theta) = 1$ e $bel(A) = 0$ para todo $A \neq \Theta$. Desde que essa função de crença parece apropriada quando não se tem evidências, ela é chamada de *função de crença vacuosa.* A classe de funções de crença pode ser caracterizada sem referenciar à atribuição de probabilidade básicas:

TEOREMA 1 *[Shafer (1976)] Se Θ é o domínio do problema, então $bel : 2^{\theta} \rightarrow [0,1]$ é uma função de crença se e somente se satisfaz às seguintes condições:*

1. $bel = (\varnothing) = 0$ *(a crença na hipótese nula é 0);*
2. $bel(\Theta) = 1$ *(a crença no domínio do problema é 1);*
3. *Para todo inteiro positivo n, e toda coleção $A_1, A_2, \cdots A_n$ de subconjuntos de Θ,*
$$bel(A_1 \cup \cdots \cup A_n) \geq \sum_{I \subseteq \{1,\cdots,n\}} (-1)^{|I|+1} bel(\bigcap_{i \in I} A_i).$$

Para a prova deste teorema é importante primeiro provar os seguintes resultados:

LEMA 1 *Se A é um conjunto finito, então*

$$\sum_{B \subseteq A} (-1)^{|B|} = \begin{cases} 1, & \text{se } A = \varnothing \\ 0, & \text{caso contrário.} \end{cases}$$

, onde $|B|$ representa o número de elementos de B.

PROVA DO LEMA 1 *O* binômio de Newton *garante que se x e a são números reais e n é um inteiro positivo, então*

$$(x+a)^n = \sum_{k=0}^{n} \binom{n}{k} a^k x^{n-k},$$

onde
$$\begin{pmatrix} n \\ k \end{pmatrix}$$
representa o número possível de combinações de n elementos tomandos k e k. Fazendo-se $x = 1$ e $a = -1$, tem-se

$$0 = (1-1)^n =$$

$$= \begin{pmatrix} n \\ 0 \end{pmatrix} (-1)^0 (1)^{n-0} - \begin{pmatrix} n \\ 1 \end{pmatrix} (-1)^1 (1)^{n-1} + \cdots + \begin{pmatrix} n \\ n \end{pmatrix} (-1)^n (1)^0 =$$

$$= \begin{pmatrix} n \\ 0 \end{pmatrix} - \begin{pmatrix} n \\ 1 \end{pmatrix} + \begin{pmatrix} n \\ 2 \end{pmatrix} - \cdots + (-1)^n \begin{pmatrix} n \\ n \end{pmatrix}.$$

- *Quando $A = \{a_1, \cdots, a_n\}$,*

$$\sum_{B \subseteq A} (-1)^{|B|} =$$

$$= (-1)^{|\varnothing|} + \sum_i (-1)^{|\{a_i\}|} + \sum_{i<j} (-1)^{|\{a_i, a_j\}|} + \sum_{i<j<k} (-1)^{|\{a_i, a_j, a_k\}|} + \cdots + (-1)^{|A|}$$

$$= (-1)^{\circ} + \sum_i (-1)^1 + \sum_{i<j} (-1)^2 + \sum_{i<j<k} (-1)^3 + \cdots + (-1)^n =$$

$$= \begin{pmatrix} n \\ 0 \end{pmatrix} - \begin{pmatrix} n \\ 1 \end{pmatrix} + \begin{pmatrix} n \\ 2 \end{pmatrix} - \cdots + (-1)^n \begin{pmatrix} n \\ n \end{pmatrix} = 0,$$

  onde $i, j \in \mathbb{N}$

- *Quando $A = \varnothing$,*

$$\sum_{B \subseteq A} (-1)^{|B|} = (-1)^{|A|} = 1.$$

Então, dada uma atribuição básica de probabilidade m, a função $bel\colon 2^{\theta} \to [0,1]$ definida como $bel(A) = \sum_{X \subseteq A} m(X)$ é uma função de crença.

$\square$

PROVA DO TEOREMA 1 *Se a função de crença é dada por (I), para alguma atribuição de probabilidade básica m, então as condições (1) e (2) do Teorema 1 seguem da definição da atribuição de probabilidade básica, ou seja,*

1. $bel(\varnothing) = \sum_{B \subseteq \varnothing} m(B) = m(\varnothing) = 0;$

2. $bel(\Theta) = \sum_{B \subseteq \Theta} m(B) = 1.$

Para a prova da condição, considere $A_1, \cdots, A_n$ *uma família fixa de subconjuntos de* Θ *e seja* $I(B) = \{i | 1 \leq i \leq n; B \subseteq A_i\}$ *para cada* $B \subseteq \Theta$ *:*

$$\sum_{\substack{I \subseteq \{1, \cdots, n \\ I \neq \varnothing\}}} \left((-1)^{|I|+1} \sum_{B \subseteq A_I} m(B) \right) = \sum_{\substack{B \subseteq \Theta \\ I(B) \neq \varnothing}} \left((m(B) \sum_{\substack{I \subseteq I(B) \\ I \neq \varnothing}} (-1)^{|I|+1} \right).$$

Simplificando a última expressão, tem-se:

$$\sum_{\substack{B \subseteq \Theta \\ I(B) \neq \varnothing}} \left(m(B) \sum_{\substack{I \subseteq I(B) \\ I \neq \varnothing}} (-1)^{|I|+1} \right) =$$

$$= \sum_{\substack{B \subseteq \Theta \\ I(B) \neq \varnothing}} \left(m(B) \left(1^{|\varnothing|} - 1^{|\varnothing|} + \sum_{\substack{I \subseteq I(B) \\ I \neq \Theta}} (-1)^{|I|+1} \right) \right) =$$

$$= \sum_{\substack{B \subseteq \Theta \\ I(B) \neq \varnothing}} \left(m(B) \left(1 + \sum_{I \subseteq I(B)} (-1)^{|I|+1} \right) \right) =$$

$$= \sum_{\substack{B \subseteq \Theta \\ I(B) \neq \varnothing}} \left(m(B) \left(1 + \sum_{I \subseteq I(B)} (-1)^{|I|} (-1)^1 \right) \right) =$$

$$= \sum_{\substack{B \subseteq \Theta \\ I(B) \neq \varnothing}} \left(m(B) \left(1 - \sum_{I \subseteq I(B)} (-1)^{|I|} \right) \right).$$

Utilizando o Lema 1 e sabendo que $I(B) \neq \varnothing$, *tem-se que:*

$$\sum_{I \subseteq I(B)} (-1)^{|I|} = 0,$$

donde

$$\sum_{\substack{B\subseteq\Theta\\ I(B)\neq\varnothing}}\left(m(B)\left(1-\sum_{I\subseteq I(B)}(-1)^{|I|}\right)\right)=\sum_{\substack{B\subseteq\Theta\\ I(B)\neq\varnothing}}(m(B)(1-0))=$$

$$\sum_{\substack{B\subseteq\Theta\\ I(B)\neq\varnothing}}m(B)=bel(A_i\cup\cdots\cup A_n).$$

Dado $A\subseteq\Theta$, $bel(A)+bel(\overline{A})\leq 1$, *pois:*

$$1=bel(\Theta)=bel(A\cup\overline{A})\geq bel(A)+bel(\overline{A})-bel(A\cap\overline{A}),$$

uma vez que $A\cap\overline{A}=\varnothing, bel(A\cap\overline{A})=0$*.*

Logo, dada uma função $\phi:2^{\theta}\rightarrow[0,1]$*, com* $\phi(\varnothing)=0$ *e* $\phi(\Theta)=1$*, uma condição necessária mais não suficiente para que* ϕ *seja uma função de crença é que* $\phi(A)+\phi(\overline{A})\leq 1$ *para todo* $A\subseteq\Theta$*.*

□

É importante notar que bel e m têm o mesmo valor em cada uma das hipótese primitivas e que bel é maior ou igual a m em conjuntos que contêm mais do que um elemento, ou seja, se $A\subseteq\Theta$ não for uma hipótese primitiva, então $bel(A)$ é a soma dos valores de m para todo subconjunto na árvore que tem A por raiz, como por exemplo:

$$bel(\{s\})=m(\{s\})$$

e

$$bel(\{a,s\})=m(\{a,s\})+m(\{a\})+m(\{s\})\geq m(\{a,s\}).$$

Dada uma função de crença bel, é possível encontrar sua atribuição de probabilidade básica, como garante o Teorema 2:

TEOREMA 2 *Suponha* $bel:2^{\theta}\rightarrow[0,1]$ *é a função de crença dada pela atribuição de probabilidade básica* $m:2^{\theta}\rightarrow[0,1]$*. Então, para todo* $A\subseteq\Theta$ *:*

$$m(A)=\sum_{B\subseteq A}(-1)^{|A-B|}bel(B).$$

Para o entendimento da prova do Teorema 2 são necessários o entendimento de dois lemas:

LEMA 2 *Se A é um conjunto finito e $B \subseteq A$, então*

$$\sum_{\substack{C \\ B\subseteq C\subseteq A}} (-1)^{|C|} = \begin{cases} (-1)^{|C|}, & \textit{se } A = B \\ 0, & \textit{caso contrário}. \end{cases}$$

PROVA DO LEMA 2

$$\sum_{\substack{C \\ B\subseteq C\subseteq A}} (-1)^{|C|} = \sum_{D\subseteq(A-B)} (-1)^{|B\cup D|} =$$

$$= \sum_{D\subseteq(A-B)} (-1)^{|B|}(-1)^{|D|} = (-1)^{|B|} \sum_{D\subseteq(A-B)} (-1)^{|D|}.$$

Tem-se duas situações:

1. *$A = B$, o que, pelo Lema 1 garante-se que:*

$$(-1)^{|B|} \sum_{D\subseteq(A-B)} (-1)^{|D|} = (-1)^{|B|}1 = (-1)^{|B|} = (-1)^{|A|},\ pois A-B = \varnothing$$

2. *$A \neq B$, o que, pelo Lema 1 garante-se que:*

$$(-1)^{|B|} \sum_{D\subseteq(A-B)} (-1)^{|D|} = (-1)^{|B|}0 = 0,\ pois A - B = \varnothing.$$

□

LEMA 3 *Seja Θ um conjunto finito e f e g funções em 2^{θ}. Então:*

$$f(A) = \sum_{B\subseteq A} g(B), \quad para \quad todo \quad A \subseteq \Theta,$$

se e somente se

$$g(A) = \sum_{B\subseteq A} (-1)^{|A-B|} f(B), \quad para \quad todo \quad A \subseteq \Theta.$$

PROVA DO LEMA 3 *Segue do Lema 2. Se $f(A) = \sum_{B\subseteq A} g(B)$ para todo $A \subseteq \Theta$, então*

$$\sum_{B\subseteq A} (-1)^{|A-B|} f(B) = (-1)^{|A|} \sum_{B\subseteq A} (-1)^{|B|} f(B) = (-1)^{|A|} =$$

$$= \sum_{B \subseteq A} \left((-1)^{|B|} \sum_{C \subseteq B} g(C) \right).$$

Mas,

$$(-1)^{|A|} \sum_{B \subseteq A} \left((-1)^{|B|} \sum_{C \subseteq B} g(C) \right) = (-1)^{|A|} \sum_{C \subseteq A} \left(g(C) \sum_{\substack{B \\ C \subseteq B \subseteq A}} (-1)^{|B|} \right)$$

com simples rearranjo dos termos. Utilizando-se do Lema 2, tem-se

$$(-1)^{|A|} \sum_{C \subseteq A} \left(g(C) \sum_{\substack{B \\ C \subseteq B \subseteq A}} (-1)^{|B|} \right) (-1)^{|A|} g(A) (-1)^{|A|} = g(A),$$

para todo $A \subseteq \Theta$.

Por sua vez, se

$$g(A) = \sum_{B \subseteq A} (-1)^{|A-B|} f(B),$$

para todo $A \subseteq \Theta$, *então:*

$$\sum_{B \subseteq A} g(B) = \sum_{B \subseteq A} \left(\sum_{C \subseteq B} (-1)^{|B-C|} f(C) \right) =$$

$$= \sum_{C \subseteq B} \left((-1)^{|C|} f(C) \sum_{\substack{B \\ C \subseteq B \subseteq A}} (-1)^{|B|} \right) =$$

$$(-1)^{|A|} f(A) (-1)^{|A|} = f(A) \quad para \quad todo \quad A \subseteq \Theta.$$

□

PROVA DO TEOREMA 2 *Seja* $bel : 2^\theta \rightarrow [0,1]$ *é a função de crença dada pela atribuição de probabilidade básica* $m : 2^\theta \rightarrow [0,1]$*:*

$$bel(A) = \sum_{X \subseteq A} m(X), \quad para \quad todo \quad A \subseteq \Theta.$$

Aplicando-se diretamente o Lema 3, tem-se:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} bel(B), \quad para \quad todo \quad A \subseteq \Theta.$$

□

Na TDS, o principal interesse são aqueles subconjuntos de Θ que têm atribuição de probabilidade básica não nula. Cada um desses subconjuntos é chamado de *elemento focal* da função de crença bel sobre 2^θ. A união de todos os elementos focais, para uma função de crença, é chamada de *núcleo*, também chamado de *centro*. Suponha que $m(\{s\}) = 0.2, m(\{a\}) = 0.3, m(\{r\}) = 0.1, m(\{s,a\}) = 0.4$, o núcleo da função de crença é $\{s,a,r\}$. Além disso, $bel(\{s,a\}) = m(\{s\}) + m(\{a\}) + m(\{s,a\}) = 0.9$. Se C for o núcleo de uma função de crença bel sobre Θ, então $B \subseteq \Theta$ satisfaz $bel(B) = 1$ se e somente se $C \subseteq B$.

3.3 Número e Função de Comunalidade

Considerando os elementos de Θ sendo pontos, dado $A \subseteq \Theta$, pode ser de interesse representar a massa de probabilidade total que pode ser movida para os pontos de A. A essa quantia dá-se o nome de *número de comunalidade* de A, representada por $Q(A)$. Já à função que calcula o número de comunalidade para todo $A \subseteq \Theta$ dá-se o nome de *função de comunalidade*. Pela definição, tem-se então que a função de comunalidade é uma função $Q : 2^\theta \to [0,1]$ tal que;

$$Q(A) = \sum_{\substack{B \subseteq \Theta \\ A \subseteq B}} m(B).$$

A comunalidade de A é a soma das atribuições de todos os conjuntos que contêm A. Representa, dessa forma, a quantidade de crença que pode ser refinada até A.

EXEMPLO 9 *Sejam* $A = \{s\}$, $B = \{a\}$ *e* $C = \{s,a\}$ *e* $m\{s\} = 0.2$, $m\{a\} = 0.3$, $m\{r\} = 0.1$ *e* $m\{a,s\} = 0.4$. *Neste caso, a atribuição de probabilidade básica em* A *e* B *poderia ser acrescida com o refinamento da atribuição básica em* C. *Isso pode ser expresso pelas comunalidades de* A *e* B*, maiores que as respectivas atribuições de probabilidade básica:*

$$Q(A) = \sum_{\substack{X \subseteq \Theta \\ A \subseteq X}} m(X) = m(A) + m(C) = 0.2 + 0.4 = 0.6$$

e

$$Q(B) = \sum_{\substack{X\subseteq\Theta \\ B\subseteq X}} m(X) = m(B) + m(C) = 0.3 + 0.4 = 0.7.$$

Facilmente percebe-se que $Q(\varnothing) = 1$. Além disso, função de crença pode ser expressa através da função de comunalidade, e vice-versa, como garante o Teorema 3. Antes de sua demonstração, entretanto dois lemas são necessários:

LEMA 4 *Seja Θ e sejam f e g funções em 2^{θ}. Então:*

$$f(A) \sum_{B\subseteq A} (-1)^{|B|+1} g(B), \quad para \quad todo \quad A \subseteq \Theta$$

se e somente se

$$g(A) = \sum_{B\subseteq A} (-1)^{|B|+1} f(B), \quad para \quad todo \quad A \subseteq \Theta.$$

PROVA DO LEMA 4 *Se*

$$f(A) = \sum_{B\subseteq A} (-1)^{|B|+1} g(B)$$

para todo $A \subseteq \Theta$, então

$$\sum_{B\subseteq A} (-1)^{|B|+1} f(B) = \sum_{B\subseteq A} \left((-1)^{|B|+1} \sum_{C\subseteq B} (-1)^{|B|+1} g(C) \right) =$$

$$= \sum_{C\subseteq A} \left((-1)^{|C|} g(C) \sum_{C\subseteq B\subseteq A} (-1)^{|B|} \right).$$

Utilizando o Lema 2, tem-se

$$\sum_{C\subseteq A} \left((-1)^{|C|} g(C) \sum_{\substack{B \\ C\subseteq B\subseteq A}} (-1)^{|B|} \right) = g(A), \quad para \quad todo \quad A \subseteq \Theta.$$

□

LEMA 5 *Seja* Θ *e sejam* f *e* g *funções em* 2^{θ}. *Então:*

$$f(A) = \sum_{B \subseteq \overline{A}} (-1)^{|B|} g(B), \quad para \quad todo \quad A \subseteq \Theta,$$

se e somente se

$$g(A) = \sum_{B \subseteq A} (-1)^{|B|} f(\overline{B}).$$

PROVA DO LEMA 5 *Seja* $h(A) = -f(\overline{A})$ *para todo* $A \subseteq \Theta$. *Se*

$$f(A) = \sum_{B \subseteq \overline{A}} (-1)^{|B|} g(B),$$

para todo $A \subseteq \Theta$, *então:*

$$h(A) = -f(\overline{A}) = -\sum_{B \subseteq \overline{A}} (-1)^{|B|} g(B) =$$

$$= \sum_{B \subseteq A} (-1)^{|B|} (-1)^{1} g(B) = \sum_{B \subseteq A} (-1)^{|B|+1} g(B).$$

Pelo Lema 4,

$$g(A) = \sum_{B \subseteq A} (-1)^{|B|+1} h(B) = \sum_{B \subseteq A} (-1)^{|B|+1} (-f(\overline{B})) =$$

$$= \sum_{B \subseteq A} (-1)^{|B|} (-1)^{1} (-1)^{1} f(\overline{B}) =$$

$$\sum_{B \subseteq A} (-1)^{|B|} f(\overline{B}).$$

Por sua vez, se

$$g(A) = \sum_{B \subseteq A} (-1)^{|B|+1} f(\overline{B}),$$

então $\sum_{B \subseteq \overline{A}} (-1)^{|B|} g(B) = -h(\overline{A}) = f(A)$.

□

TEOREMA 3 *Seja* $Q : 2^{\Theta} \to [0,1]$*, uma função de comunalidade, então* $bel : 2^{\Theta} \to [0,1]$ *dada por*

$$bel(A) = \sum_{B \subseteq \overline{A}} (-1)^{|B|} Q(B)$$

é a função de crença associada a essa função de comunalidade. Por sua vez, se $bel : 2^{\Theta} \to [0,1]$ *é uma função de crença, então* $Q(A) = \sum_{B \subseteq A} (-1)^{|B|} bel(\overline{B})$ *é a função de comunalidade associada a essa função de crença.*

PROVA DO TEOREMA 3 *Seja* $Q : 2^{\theta} \to [0,1]$*, uma função de comunalidade, então* $bel : 2^{\theta} \to [0,1]$ *dada por* $bel(A) = \sum_{B \subseteq \overline{A}} (-1)^{|B|} Q(B)$ *é a função de crença associada a essa função:*

$$\sum_{B \subseteq \overline{A}} (-1)^{|B|} Q(B) =$$

$$\sum_{B \subseteq \overline{A}} \left((-1)^{B} \sum_{\substack{C \\ B \subseteq C}} m(C) \right) = \sum_{C \subseteq \Theta} \left(m(C) \sum_{B \subseteq C \cap \overline{A}} (-1)^{|B|} \right).$$

Pelo Lema 1,

$$\sum_{C \subseteq \Theta} \left(m(C) \sum_{B \subseteq C \cap \overline{A}} (-1)^{|B|} \right) = \sum_{\substack{C \\ C \cap \overline{A} = \varnothing}} m(C) = \sum_{C \subseteq A} m(A) = bel(A),$$

para todo $A \subseteq \Theta$*. Pelo Lema 5, obtém-se diretamente*

$$Q(A) = \sum_{B \subseteq A} (-1)^{|B|} bel(\overline{B}).$$

□

Outra propriedade importante é dada pelo Teorema 4:

TEOREMA 4 *Seja* C *o centro de uma função de crença sobre* Θ *e* Q *sua função de comunalidade. Então um elemento* $\theta \in \Theta$ *está em* C *se e somente se* $Q(\{\theta\}) \geq 0$.

PROVA DO TEOREMA 4 *A equação* $Q(\{\theta\}) = \sum_{\substack{B \subseteq \Theta \\ \theta \in B}} m(B), Q(\{\theta\})$ *será positiva se e somente se for um elemento focal, isto é,se estiver em* C.

$\square$

Uma vez que a função de comunalidade é não incremental segue da conclusão que $Q(A) = 0$ quando A inclui um ponto fora do centro C.

Ainda da relação entre função de crença e função de comunalidade, tem-se:

$$bel(\varnothing) = 0 = \sum_{B \subseteq \Theta} (-1)^{|B|} Q(B) \quad ou \quad \sum_{\substack{A \subseteq \Theta \\ A \neq \varnothing}} (-1)^{|A|+1} Q(A) = 1.$$

3.4 Plausibilidade de uma Função de Crença

Dado $A \subseteq \Theta$, o valor $bel(A)$ pode não evidenciar totalmente o quanto se pode acreditar em $\overline{A}$. Uma descrição mais completa pode ser dada pelo *grau de dúvida*, denominado *dou*, definido por

$$dou(A) = bel(\overline{A}).$$

O grau de dúvida é utilizado com menos freqüência que a quantidade $1 - bel(\overline{A})$, denominada *plausibilidade* de A, ou também *probabilidade superior* de A, notada por $\wp l(A)$, que fornece a quantidade máxima de crença que pode ser atribuída a A. Desde que $bel(A) + bel(\overline{A}) \leq 1$, tem-se que $bel(A) \leq \wp l(A)$ para todo $A \subseteq \Theta$. Pode-se também expressar $\wp l(A)$ em termos da atribuição básica de probabilidade m como garante o Teorema 5:

TEOREMA 5 *Dada uma função de crença $bel : 2^{\Theta} \to [0,1]$ e sua atribuição de probabilidade básica $m : 2^{\Theta} \to [0,1]$, a função $\wp ll : 2^{\Theta} \to [0,1]$ dada por*

$$\wp l(A) = \sum_{B \cap A \neq \varnothing} m(B)$$

é a função de plausibilidade.

PROVA DO TEOREMA 5 *Para qualquer $A \subseteq \Theta$, tem-se*

$$\wp l(A) = 1 - bel(\overline{A}) \sum_{B \subseteq \Theta} m(B) - \sum_{B \subseteq \overline{A}} (B) = \sum_{B \cap (A) \neq \varnothing} m(B).$$

Da definição de plausibilidade acima e da definição da função de comunalidade, dada por

$$Q(A) = \sum_{\substack{X \subseteq \Theta \\ A \subseteq B}} m(B),$$

tem-se que, para todo elemento particular θ de Θ,

$$\wp l(\{\theta\}) = Q(\{\theta\}).$$

□

3.5 Intervalos de Crença

Como visto anteriormente, a plausibilidade de uma da hipótese A, $\wp l(A)$, representa o quanto é possível acreditar em A. Se $bel(A)$ representa a crença atual em A, e sabendo-se que $bel(A) \leq \wp l(A)$, é natural que a informação contida na crença em A seja mais convenientemente expressa pelo intervalo $[bel(A), \wp l(A)]$ ao invés de $bel(A)$ apenas.

Com as funções de crença Bayesianas, utilizadas pela teoria de probabilidade clássica, ocorre que $bel(A) = \wp l(A)$, resultando que o intervalo $[bel(A), \wp l(A)]$ é degenerado, ou seja, possui um único ponto. Em geral, entretanto, isso não ocorre na TDS. É desejável, portanto, que sistemas baseados na TDS ao fornecerem informações de crença em uma dada hipótese ou evidência forneçam não somente o grau de crença, mas o intervalo $[bel(A), \wp l(A)]$ que expressa a faixa de valores no qual é possível acreditar em A, sem incorrer em erros graves de suposição. Esse intervalo recebe apropriadamente o nome de *intervalo de crença*, representado por $\Im(A)$, e é tão mais amplo quanto mais incerteza houver sobre a crença em A. Isso pode ser visualizado com clareza na função de crença vacuosa, onde todas as hipóteses primitivas possuem $[0, 1]$ como intervalo de crença.

EXEMPLO 10 *Sejam $A = \{s\}, B = \{a\}$ e $C = \{s, a\}$ como no Exemplo 2, onde $\Theta = \{a, i, s, r\}$. Neste caso, os intervalos de crença dessas hipóteses são:*

$$\Im(A) = [bel(A), \wp l(A)] = [bel(A), 1 - bel(\overline{A})] = [bel(\{s\}), 1 - bel(\{a, i, r\})] =$$

$$= [0.2, 1 - (0.3 + 0.1)] = [0.2, 0.6]$$

$$\Im(B) = [bel(B), \wp l(B)] = [bel(B), 1 - bel(\overline{B})] = [bel(\{a\}), 1 - bel(\{s, i, r\})] =$$

$$= [0.3, 1 - (0.2 + 0.1)] = [0.3, 0.7]$$

$$\Im(C) = [bel(C), \wp l(C)] = [bel(C), 1 - bel(\overline{C})] = [bel(\{s, a\}), 1 - bel(\{i, r\})] =$$

$$= [(0.2 + 0.3 + 0.4), 1 - (0.1)] = [0.9, 0.9]$$

A crença tanto em A quanto em B pode ser aumentada em 0.4 pontos, o que em ambas representam um aumento maior que o dobro do grau de crença. A crença em C, entretanto está em seu máximo, não sendo possível nenhum acréscimo, representando uma certeza de que a probabilidade de C ocorrer é 0.9.

3.6 Funções de Crença Bayesianas

Na literatura podem ser encontradas várias referências que abordam o relacionamento entre a TDS e a Teoria de Bayes [Pearl (1982)]. É fato que o conceito de função de crença é suficientemente amplo, a ponto de permitir que os conceitos da Teoria Bayesiana possam ser focalizados sob a perspectiva da TDS. Mais ainda, as probabilidades Bayesianas podem ser enquadradas como casos específicos das funções de crença. São as *funções de crença Bayesianas*.

Uma função de crença *bel* é dita bayesiana se $bel(A \cup B) = bel(A) + bel(B)$ para todo $A, B \subseteq \Theta$ e $A \cap B = \varnothing$. Em síntese, uma função de crença Bayesiana usa a atribuição de probabilidade básica $m : 2^{\Theta} \rightarrow [0, 1]$ tal que:

- $m(\{\theta\}) = bel(\{\theta\})$ para todo $\theta \in \Theta$;
- $m(A) = 0$ para todo $A \subseteq \Theta$ que não seja conjunto unitário.

O próximo teorema prova várias equivalências envolvendo funções de crença bayesinas.

TEOREMA 6 *Seja uma função de crença* $bel : 2^{\Theta} \rightarrow [0, 1]$ *com plausibilidade* $\wp l(A) : 2^{\Theta} \rightarrow [0, 1]$ *e comunalidade* $Q : 2^{\Theta} \rightarrow [0, 1]$. *As seguintes assertivas são todas equivalentes entre si:*

1. bel *é bayesiana;*
2. *os elementos focais de* bel *são conjuntos unitários;*
3. bel *garante comunalidade zero para qualquer subconjunto contendo mais que um elemento;*
4. $bel(A) = \wp l(A)$ *para todo* $A \subseteq \Theta$*;*
5. $bel(A) + bel(\overline{A}) = 1$ *para todo* $A \subseteq \Theta$.

PROVA DO TEOREMA 6 *Tem-se*

- $(1) \equiv (2)$: *Segue diretamente do fato que a atribuição de probabilidade básica é nula para conjuntos não unitários.*

- $(2) \equiv (3)$: *Segue da relação*

$$Q(A) = \sum_{\substack{X \subseteq \Theta \\ A \subseteq B}} m(B),$$

lembrando que se B possui mais que um elemento, então $m(B) = 0$.

- $(2) \equiv (4)$: *Segue da comparação entre as relações*

$$bel(A) = \sum_{B \subseteq A} m(B) \quad e \quad \wp l(A) = \sum_{B \cap A \neq \varnothing} m(B),$$

lembrando que se B *possui mais que um elemento, então* $m(B) = 0$.

- $(4) \equiv (5)$: *Segue da relação que* $\wp l(A) = 1 - bel(A)$.

□

Uma outra importante propriedade das funções de crença bayesianas é mostrado no Teorema 7 que se segue:

TEOREMA 7 *Uma função* $bel : 2^{\Theta} \to [0, 1]$ *é uma função de crença bayesianas se existe uma função* $\rho : \Theta \to [0, 1]$ *tal que:*

$$\sum_{\theta \in \Theta} \rho(\theta) = 1 \quad e \quad bel(A) = \sum_{\theta \in A} \rho(\theta).$$

PROVA DO TEOREMA 7 *Tem-se*

$$bel(\varnothing) = \sum_{\theta \in \varnothing} \rho(\theta) = 0,$$

$$bel(\Theta) = \sum_{\theta \in \Theta} \rho(\theta) = 1.$$

Quando $A \cap B = \varnothing, A, B \subseteq \Theta$

$$bel(A) + bel(B) = \sum_{\theta \in A} \rho(\theta) + \sum_{\theta \in B} \rho(\theta) = \sum_{\theta \in A \cup B} \rho(\theta) = bel(A \cup B).$$

Pode-se verificar que a função $\rho : \Theta \to [0, 1]$ *satisfaz* $\rho(\theta) = m(\{\theta\})$.

□

3.7 Combinação da Função de Crença

A TDS combina diferentes funções de crença para fazer a propagação de crença. Para isto é necessário o cálculo da soma ortogonal das funções de crença utilizando a regra de combinação de Dempster. Esta combinação é importante por exemplo, em um diagnóstico médico porque requer combinações que suporte uma hipótese, ou à sua negação, com base no acúmulo de múltiplas observações.

Para indicar os efeitos combinados de duas atribuições de probabilidades básica m_1 e m_2 é utilizada a notação $m_1 \oplus m_2$. A partir desta notação, a função de crença correspondente, notada por $bel_1 \oplus bel_2$, pode ser calculada facilmente.

O cálculo para a soma ortogonal de duas atribuições m_1 e m_2 em um domínio do problema Θ é:

$$m(\varnothing) = 0,$$

$$m_1 \oplus m_2(A) = \chi \sum_{\substack{X \cap Y = A \\ A \neq \varnothing}} m_1(X) \times m_2(Y), \;\; para \;\; todo \;\; A \subseteq \Theta,$$

onde χ é a constante de normalização, definida como $\frac{1}{1-\kappa}$, e κ é a soma dos bpa's de todas as ocorrências do conjunto $\varnothing$. O conjunto vazio ocorre quando se tenta combinar hipóteses disjuntas, ou seja indica que existem evidências que suportam hipótese que estão em conflito, uma com a outra.

A agregação de evidências deve ser independente da ordem na qual ela acontece, ou seja a soma ortogonal produz o mesmo valor independentemente da ordem na qual as funções são combinadas. Para ser possível a combinação de crença utilizando a regra de combinação de Dempster é preciso que a função que se segue seja válida:

$$\sum_{A_i \cap B_i = \varnothing} m_i(A_i) m_2(B_j) < 1.$$

Também pode ser verificado que, se Q_1 e Q_2 são as funções de comunalidade associadas a bel_1 e bel_2, então a combinação das funções de comunalidade, denotada por $Q_1 \oplus Q_2$, é dada por $Q_1 \oplus Q_1(A) = \chi Q_1(A) Q_2(A)$ para qualquer não vazio $A \subseteq \Theta$. Para melhor entendimento segue os Exemplos 11 e 12.

EXEMPLO 11 *Considere novamente o frame de discernimento*

$$\Theta = \{a, i, r, s\}$$

e suponha que, para um certo paciente, um determinada observação m_1 *indique* $\{a, i\}$*, com grau* 0.5*, enquanto que uma outra* m_2 *desconfirme alergia com grau*

0.6. *A rede de crença, baseada em ambas observações, é dada por* $m_1 \oplus m_2$ *e é representada na Tabela 3.3:*

Tabela 3.3: Rede de Crenças para $m_1 \oplus m_2$

		$\{i,s,r\}$	(0.6)	Θ	(0.4)
$\{a,i\}$	(0.5)	$\{i\}$	(0.3)	$\{a,i\}$	(0.2)
Θ	(0.5)	$\{i,r,s\}$	(0.3)	Θ	(0.2)

Então,

$$m_1 \oplus m_2(\{i\}) = 0.30,$$

$$m_1 \oplus m_2(\{i,r,s\}) = 0.30,$$

$$m_1 \oplus m_2(\{a,i\}) = 0.20,$$

$$m_1 \oplus m_2(\{\Theta\}) = 0.20,$$

$$m_1 \oplus m_2 = 0 \ \ para \ \ quaisquer \ \ outros \ \ subconjuntos \ \ de \ \ \Theta.$$

Note que nesse exemplo $\chi = 1$ uma vez que $\kappa = 0$

A partir da Tabela 3.3 pode-se então calcular $bel_1 \oplus bel_2$, para todos os elementos de 2^{Θ}. Por exemplo,

$$bel_1 \oplus bel_2(\{a,i\}) =$$

$$= m_1 \oplus m_2(\{a,i\}) + m_1 \oplus m_2(\{a\}) + m_1 \oplus m_2(\{i\}) =$$

$$= 0.2 + 0 + 0.3 = 0.5.$$

EXEMPLO 12 *Suponha agora que para o mesmo paciente do exemplo anterior, uma terceira evidência confirma o diagnóstico de alergia com grau de 0.8, como ilustrado na Tabela 3.4. Pela TDS deve-se agora calcular, onde* $m_4 = m_1 \oplus m_2$ *do exemplo anterior*

Tabela 3.4: Rede de Crenças para $m_3 \oplus m_4$

		$\{i\}$	(0.3)	$\{a,i\}$	(0.2)	$\{i,r,s\}$	(0.3)	Θ	(0.2)
$\{a\}$	(0.8)	$\varnothing$	(0.24)	$\{a\}$	(0.16)	$\varnothing$	(0.24)	$\{a\}$	(0.16)
Θ	(0.2)	$\{i\}$	(0.06)	$\{a,i\}$	(0.04)	$\{i,r,s\}$	(0.06)	Θ	(0.04)

Como neste exemplo o conjunto $\varnothing$ foi obtido duas vezes com valor de crença 0.24,

$$\kappa = 0.24 + 0.24 = 0.48 \quad e \quad 1 - \kappa = 0.52.$$

Então,

$$m_3 \oplus m_4(\{a\}) = (0.16 + 0.16)/0.52 = 0.615,$$

$$m_3 \oplus m_4(\{i\}) = 0.06/0.52 = 0.115,$$

$$m_3 \oplus m_4(\{i, r, s\}) = 0.06/0.52 = 0.115,$$

$$m_3 \oplus m_4(\{a, i\}) = 0.04/0.52 = 0.077,$$

$$m_3 \oplus m_4(\{\Theta\}) = 0.04/0.52 = 0.077,$$

$$m_3 \oplus m_4 = 0 \ para \ quaisquer \ outros \ subconjuntos \ de \ \Theta.$$

Note que nesse exemplo $\sum m_3 \oplus m_4 = 1$, como requer a definição de um bpa.

Em [Yager et al.(1994)] encontram-se contribuições para a combinação de crença, bem como os avanços da TDS.

3.8 Peso de Conflito

Com as possíveis combinações de funções de crença, alguns problemas podem surgir.

Como definido anteriormente, $\chi = \frac{1}{1-\kappa}$, onde κ é a soma dos bpa's de todas as ocorrências do conjunto $\varnothing$, cujas ocorrências se devem à combinação de hipóteses disjuntas. O valor $\log(\chi)$ é denominado *peso de conflito* entre bel_1 e bel_2 denotado *con*(bel_1, bel_2). Se bel_1 e bel_2 não se conflitam em nada, então $\kappa = 0$ e $con(bel_1, bel_2) = 0$. Por outro lado, se bel_1 e bel_2 não possuem nenhuma evidência em comum, ou seja a união de todas as intersecções de suas evidências é o conjunto vazio (hipótese nula), então $\kappa = 1$ e $con(bel_1, bel_2) = \infty$. Em tais condições, a combinação $bel_1 \oplus bel_2$ não é possível.

Em alguns casos, mesmo sendo possível a combinação de duas funções de crença utilizando-se da regra de Dempster, o resultado pode não ser o esperado, sendo até mesmo contrário à intuição. Nesses casos, entretanto, o peso de conflito fornece dados fundamentais sobre a combinação dessas crenças. Pode ser verificado que quanto maior o peso de conflito, mais indesejável será o resultado.

EXEMPLO 13 *: Suponha que dois doutores examinem um paciente e concordam que ele sofre ou de meningite (M), ou de convulsão (C) ou de tumor cerebral (T). Ou seja,* $\Theta = \{M, C, T\}$*. No entanto discordam quanto ao diagnóstico. Para o primeiro médico as atribuições de probabilidades básicas deveriam ser:*

$$m_1(\{M\}) = 0.9 \qquad e \qquad m_1(\{T\}) = 0.1,$$

enquanto para os segundo seriam:

$$m_2(\{C\}) = 0.8 \qquad e \qquad m_2(\{T\}) = 0.2,$$

A rede de crenças que combina essas duas atribuições é dada pela Tabela 3.5:

Tabela 3.5: Rede de Crenças para $m_1 \oplus m_2$

		$\{M\}$	(0.9)	$\{T\}$	(0.1)
$\{C\}$	(0.8)	$\varnothing$	(0.72)	$\varnothing$	(0.08)
$\{T\}$	(0.2)	$\varnothing$	(0.18)	$\{T\}$	(0.02)

Neste caso, $\kappa = 0.72 + 0.08 + 0.18 = 0.98$ *o que dá*

$$\chi = \frac{1}{1-\kappa} = \frac{1}{0.02} = 50 \quad e \quad con(bel_1, bel_2) = \log(50) \cong 1.698.$$

Observe que, temos $m_1 \oplus m_2(T) = 1$*, o que não é um resultado esperado pela intuição.*

De forma geral, quando o κ for muito maior que a metade, não é desejável a combinação de duas funções de crença. Então impondo um valor limite para κ é possível determinar um máximo para o peso de conflito.

Alguns autores sugerem novas formas de combinação que não seja a regra de Dempster. Walley em [Walley (1996)] sugere que sejam combinadas por extensão natural e Wang em [Wang (1994)] sugere uma nova abordagem para a TDS, utilizando freqüências inferiores e superiores. Além disso, em [Yager et al.(1994)] são apresentadas algumas maneiras de combinar crenças. Como estas abordagem utilizam conceitos avançados e complexos, foge aos objetivos deste trabalho.

Capítulo 4

O Uso da Teoria de Dempster-Shafer para Representação de Incerteza em Sistemas Baseados em Conhecimento

4.1 Avaliando a Teoria de Dempster-Shafer

Walley em [Walley (1996)] discute a necessidade do estabelecimento de critérios para a avaliação das medidas de incerteza e propõe seis critérios, básicos:

1. **Interpretação:** a medida deve ter uma interpretação clara que seja suficientemente precisa para:
   - ser usada;
   - entender as conclusões do sistema e usar tais conclusões para deflagrar as ações correspondentes;
   - estabelecer regras para a combinação de tais medidas e para a sua atualização.
2. **Imprecisão:** a medida deve ser capaz de modelar ignorância parcial ou incompleta, informação limitada ou conflitante, bem como declarações impre-

cisas de incerteza. É importante lembrar que ignorância parcial e informações conflitantes são comuns em domínios reais.

3. **Consistência:** por um lado, o sistema de tratamento de incerteza deve fornecer métodos que permitam a verificação da consistência de todas as declarações de incerteza e de todas as suposição *default.* Por outro lado, as regras de cálculo devem garantir que as conclusões sejam consistentes com todas as declarações e suposição *defaults.*

4. **Cálculo:** devem existir regras para combinar, atualizar e usar medidas de incertezas.

5. **Declaração (Input):** o sistema deve cuidar para que o usuário não tenha problemas quando do fornecimento de todos os valores de incerteza necessários como entrada do sistema. Além disso, um sistema para o tratamento de incertezas deve viabilizar a combinação de avaliações qualitativas com valores quantitativos de incerteza.

6. **Computação:** deve ser computacionalmente factível para o sistema derivar inferências e conclusões a partir das declarações iniciais.

Walley classifica esses seis critérios em teóricos e práticos. Os teóricos (um, dois, três, quatro), são aqueles que, para serem verificados necessitam ser subsidiados por uma teoria adequada de incerteza que viabilize tal verificação, independentemente do domínio da aplicação. Já os práticos (cinco, seis), dependendo da aplicação, podem ou não serem satisfeitos. São dependentes do tipo de modelo utilizado, do número de entradas necessárias, da restrição de tempo, do poder computacional e da habilidade do usuário.

É possível fazer uma análise da TDS de acordo com os seis critérios.

- **Interpretação:** As funções de crença permitem rápida compreensão, o que facilita o fornecimento de informação ao usuário. sua estrutura é semelhante as de Fatores de Certeza sob certos aspectos. A TDS teria vantagens sobre os Fatores de Certeza por possuir uma rigorosa estrutura matemática. Quando expressa por intervalos de crença, os resultados são ainda melhores.

- **Imprecisão:** A TDS permite a representação de ignorância parcial e total, e conflito. No entanto, deve-se ter um cuidado especial na combinação de crenças altamente conflitantes, uma vez que isso pode levar a resultados contra-intuitivos.

- **Cálculo:** A regra de Dempster permite a combinação de funções de crença, desde que essas não sejam muito conflitantes. Na maioria das situações a regra de Dempster produz resultados satisfatórios.

- **Consistência:** O peso de conflito permite verificar se a combinação de duas ou mais crenças produzirá resultados indesejáveis.

- **Declaração:** Não há ainda procedimentos seguros que guiem o processo de declaração de crenças, o que pode introduzir agravantes ao sistema.

- **Computação:** Como a TDS utiliza um grande número de hipótese, o cálculo computacional pode ser extremamente penoso. No entanto, por não precisar de probabilidades condicionais e como uma evidência em geral foca apenas um pequeno número do total das hipóteses, o cálculo pode ser bastante simplificado, pois a TDS não precisa indicar crença nula. Outro problema em questionamento é o de como as crenças devem ser propagadas.

4.2 Comparações entre a Teoria de Dempster-Shafer e outros Modelos

Dentre as teorias apresentadas, a teoria da probabilidade subjetiva, é a mais antiga e a mais bem estabilizada, conseqüentemente a que mais atrai atenções. Entretanto, muitos pesquisadores acham os métodos bayesianos de propagação de crença ineficientes. Esta insatisfação, em troca, motiva o desenvolvimento de outras teorias mais novas.

Estas novas teorias aperfeiçoaram a teoria da probabilidade subjetiva em alguma área específica: teoria de Dempster-Shafer representa melhor a ignorância; teoria da possibilidade é o melhor meio de se trabalhar dados nebulosos e fatores de certeza são melhores para explicarem o fluxo de controle de sistemas baseados em regras.

Como apontado em [Ng & Abramsom (1990)], nenhuma delas está livre de falhas ou de argumentos que as deixe inutilizável. Tanto a teoria de Dempster-Shafer, quanto a teoria da possibilidade utilizam terminologias pouco conhecidas, e tem uma falta de respaldo semântico e de implementações claras.

Muitos proponentes da probabilidade tem feito reivindicações bastantes corajosas sobre a universalidade desta teoria - Cheeseman [Cheeseman (1985)] e Pearl

[Pearl (1985)] defendem que a teoria da probabilidade pode ser aplicada em qualquer problema. Lindley [Lindley (1987)] disse que qualquer problema que possa ser feito com *lógica fuzzy*, funções de crença, ou outra técnica alternativa vai ser melhor feito com probabilidade. Mas este tipo de posição causa muita discórdia entre os pesquisadores.

Horvitz [Horvitz et al.(1986)] propôs um trabalho que pode ser usado para identificar as diferenças fundamentais entre metodologia probabilística e não probabilística, e comparar a performace de diferentes esquemas não probabilísticos. Wise e Henrion em [Wise & Henrion (1986)] compararam a performace de diferentes modelos (teoria da probabilidade subjetiva, teoria dos conjuntos fuzzy, fatores de certeza) usando o mesmo conjunto de regras e dados. Nestes testes, redes Bayesianas ganharam dos seus competidores enquanto conjuntos fuzzy e fatores de certeza, embora não com resultados equivalentes a Bayesianas obtiveram resultados iguais.

Heckerman [Heckerman (1986)] empiricamente comparou três métodos (teorema de Bayes, probabilidade e Dempster-Shafer) e defendeu que aproximações Bayesianas são melhores.

Como observado em [Ng & Abramsom (1990)], embora os vários esquemas de gerenciamento de incerteza são diferentes um dos outros, eles possuem alguns problemas em comum. Encontrar um especialista capaz de quantificar acuradamente informações pessoais, subjetiva e qualitativa não é tarefa fácil. É fácil verificar que os humanos são facilmente tendenciosos e, desta forma, a qualidade do conhecimento extraído dos especialistas depende extremamente do método usado para sua atribuição. Entretanto, pesquisadores de sistemas especialistas estão desprendendo o mínimo de esforço em estudos de sistemas especialistas e de técnicas apropriadas de atribuições em SBC's. Ainda, os trabalhos desenvolvidos nesta área não têm sido adequadamente notados pela comunidade de Inteligência Artificial.

Na opinião desses autores [Ng & Abramsom (1990)], a escolha do modelo de gerenciamento de incerteza é fortemente dependente de condições existentes e domínios. Se for possível ter um especialista a disposição para passar informações, então o mais apropriado é o uso da teoria da probabilidade. Na ausência de tal especialista, probabilidade pode se tornar muito arbitrário. Um sistema especialista no qual o conhecimento da base de dados é construído principalmente de material textual pode necessitar de propagação de informação usando paradigmas mais complicados.

Projetistas de sistemas especialistas enfrentam uma realidade - base de conhecimento grande não pode ser procurada depressa; o melhor para o que eles podem esperar é eficiência relativa.

Ainda na opinião de [Ng & Abramsom (1990)], a teoria da probabilidade parece ser o método mais eficiente no caso geral, embora existam instâncias em que outros métodos podem ter uma performace melhor que a teoria da probabilidade. Mesmo porque as técnicas de propagação de crença Bayesiana são de complexidade exponencial, o que a torna impraticável para grandes bancos de dados.

O Modelo de Fator de Certeza mesmo não sendo bem fundamentado matematicamente, na prática parece oferecer resultados satisfatórios. Eles têm sido incorretamente usados em sistemas especialistas porque o CF tem sido usado como um valor absoluto, ao invés de um valor incremental de probabilidade e o seu uso é causado pela normalização do MB e MD antes que sua diferença aritmética seja calculada. Um outro problema encontrado é a suposição da independência das evidências e o problema da inabilidade de distinguir entre ignorância e conflito, ambos representados por $CF = 0$.

Em resumo, todos os modelos têm suas vantagens e suas desvantagens. Porém, muitos problemas interessantes permanecem em aberto, os quais suas resoluções podem conduzir à descoberta de técnicas poderosas, gerais, e implementáveis por tratar de gerenciamento de incerteza em sistemas especialistas.

4.3 Comentários Finais

Como já observado nesse capítulo, a TDS permite uma representação prática e elegante de ignorância. Além disso, pôde ser verificado que ela engloba facilmente a Teoria Bayesiana e o Modelo de Fatores de Certeza.

Capítulo 5

Conclusões

Para muitos matemáticos e para a maioria das pessoas, “possibilidade” e “crença” são abordadas com o nome de “probabilidade”. A TDS rejeita essa unificação, conforme [Shafer (1976)]:

> muito dos graus numéricos de crença estudados aqui são possibilidades e não obedecem todas as regras obedecidas pelo acaso. (...) Possibilidades surgem quando alguém descreve um experimento aleatório ou randômico, como o rolar de um dado ou o atirar de uma moeda. O resultado de tal experimento varia randomicamente entre experimentos fisicamente independentes.

Dessa forma, os graus de possibilidade governando um experimento aleatório podem ou não coincidirem com os nossos graus de crença sobre o resultado desse experimento. Como afirma [Shafer (1976)]:

> Se conhecemos as possibilidades, então podemos seguramente adotá-las como graus de crença. Mas se nós conhecemos as possibilidades, então será uma extraordinária coincidência nossos graus de crença serem iguais a ela.

A TDS é uma teoria bem fundamentada matematicamente e além disso permite expressar ignorância parcial e total de forma extremamente adequada, ao contrário da Teoria de Bayes que expressa ignorância parcial atribuindo-se crença à negação de hipótese e ignorância total dividindo-se o total da crença entre as hipóteses presentes (eventualmente, atribuindo mais crença do que realmente possuem). Outra diferença importante entre a TDS e a Teoria de Bayes é modo como novas evidências são adicionadas ao sistema, uma vez que na Teoria de Bayes a ordem com que as evidências são apresentadas pode influir no sistema, além de que cada nova evidência é expressa como certeza. Na TDS, as funções são combinadas utilizando-se da regra de combinação de Dempster, conforme exposto na Seção 3.7. A regra de combinação de fatores de certeza, bem como a regra de combinação da teoria de Bayes são na realidade especializações da regra de combinação da TDS.

Essa regra garante que o resultado da combinação não depende da ordem da apresentação das evidências e nem necessita que estas expressem certeza. Dessa forma, como afirma Shafer em [Shafer (1976)]:

> Desde que não se requeira que expressamos nossa evidência como uma certeza, a regra de combinação de Dempster permite-nos construir descrições de raciocínio provável que são mais modestos que descrições bayesianas, mas mais fidedigno à forma humana de pensar.

Quando da implementação de sistemas que utilizam a TDS, é importante que sejam dadas devidas atenções para o intervalo de crença, bem como analisar o peso de conflito ao se combinar funções de crença, o qual garantirá maior segurança nas decisões, bem como fornecerá maior informação sobre as hipóteses e evidências em análise.

Uma outra observação a ser feita sobre a TDS é que não há uma concordância quanto ao uso da regra de Dempster para a combinação de crença. Observe que alguns trabalhos (ver, por exemplo, [Yager et al.(1994)], [Walley (1996)] e [Wang (1994)]) apresentam novos modelos de combinação de crenças na TDS. É importante que trabalhos futuros analisem essas alternativas, permitindo um melhor uso dessa teoria.

Apêndice A

Implementação

Neste capítulo, é apresentada uma implementação em linguagem C++ dos conceitos básicos da TDS. Essa implementação consiste de 5 arquivos: `ds.h`, `ds.cpp`, `setfunc.h`, `setfunc.cpp` e `teste.cpp`. Segue uma listagem do arquivo `ds.h` e do `setfunc.h`, para uma melhor compreensão da estrutura da implementação.

```
#ifndef DS_H
#define DS_H

#include <cstdlib>
#include <fstream>
#include <string>
#include <sstream>
#include <vector>
#include "setfunc.h" //funções auxiliares de operações entre conjuntos

class bpa{ //atribuições básica de probabilidade de um conjunto
 public:
  //construtores e destrutor
  bpa();
  bpa(vector<string> hyp, double val);
  ~bpa();

  // funções "get" e "set"
  inline double get_attr();
  inline void set_attr(double attr);
  inline vector<string> get_set();
  inline void set_set(vector<string> setX);

 private:
  vector<string> set; // conjunto ao qual será atribuído o bpa
  double attribution; // atribuição feita ao conjunto pelo bpa
};
```

```
class bpas{ //atribuições básica de probabilidade de todos os conjuntos
 public:
  //construtor e destrutor
  bpas();
  ~bpas();

  //funções extras
  inline void insertbpa(bpa& A); //insere a atribuição de um conjunto
  inline void insertbpa(vector<string> setX, double attr);  //idem
  inline void set_bpaDomain(double x); //atribui valor ao bpa do domínio
  inline bpa get_bpa(int i); //retorna um bpa do vetor de bpas
  inline int get_numbpas( ); //retorna número de bpas do vetor de bpas
  inline vector<string> get_set(int i); //retorna o vetor do i. bpa
  inline double get_attr(int i); //retorna o vetor do i. bpa
  inline double get_bpaDomain(); //retorna bpa do domínio
  inline void reconfigure_bpaDomain(); //recalcula o bpa do domínio
  inline void set_attr(int i, double value); //atribui um dado valor de
                          //atribuição a um elemento do vetor de bpas
  double bel_aux(vector<string> set); //calcula crença de um conjunto

 private:
  vector<bpa> vecbpas; //vetor de bpa o qual contém todos os conjuntos
                    //e suas respctivas atribuições básica.
  double bpaDomain; //atribuição básica de probabilidade do domínio
  bool reconfigure; //informa se necessário reconfigurar o bpa do domínio
};


class domain{ //domínio do problema
 private:
  vector<string> hypot;
  string description;
  vector<bpas> attributions;

 public:
  //construtor e destrutor
  domain(string fileName); // lê dados do arquivo
  ~domain();

  //funções auxiliares
  int readFromFile(string fileName); //lê dados de arquivo (*.dat)
  int saveToFile(string fileName); //salva dados em arquivo (*.dat)
  int numAttr(); //retorna número de atribuicões
  void insert_bpas(bpas temp); // insere uma atribuição

  //funções da TDS
  double bel(int attr, vector<string> set); //calcula crença de um conjunto
  double plausibility(int attr, vector<string> set); //calcula plausibilidade
  bpas combine(int i, int j); //combina duas atribuições
};
```

```
#endif
```

++++++++++++++++++++++++Arquivo setfunc.h++++++++++++++++++++++++

```
#ifndef SETFUNC
#define SETFUNC

#include <string>
#include <vector>
#include <algorithm>

//------------------------------funções auxiliares-------------------------
//calcula um vetor menos o outro
vector<string> opMinusSet(vector<string> setX, vector<string> setY);

//intercessão de dois vetores
vector<string> intersection(vector<string> setX, vector<string> setY);

//verifica se X inclui Y
bool includes(vector<string> setX, vector<string> setY);

//verifica se X = Y
bool equal(vector<string> setX, vector<string> setY);

#endif
```

Referências Bibliográficas

[Walley (1996)] Walley, P. Measures of uncertainty in expert systems. *Artificial Inteligence 83*, 1996, p. 1-58.

[Yager et al.(1994)] Yager, R.R.; Fedrizi, M. & Kacprzyk, J. *Advances in the Dempster-Shafer theory of evidence*, New York, John Wiley & Sons, 1994.

[Strat (1994)] Strat, T.M. Decision analysis using belief functions. In:Yager, R.R.; Fedrizi, M. & Kacprzyk, J. *Advances in the Dempster-Shafer theory of evidence*, New York, John Wiley & Sons, 1994.

[Bonissone (1991)] Bonissone, P. Plausible reasoning. In: Shapiro, S. C.; Eckroth, D. & Valassi, G. A. (eds.) *Encyclopedia of Artificial Intelligence*, New York, John Wiley & Sons, 1991, p.854-863.

[Ng & Abramsom (1990)] Ng, K.C. & Abramsom, B. Uncertainty management in expert systems. *IEEE Expert*, April 1990, p.29-47.

[Shortliffe & Buchanan (1975)] Shortliffe, E.H. & Buchanan, B.G. A model of Inexact Reasoning in Medicine. *Math. Biosci*, 23, 1975, p. 351-379.

[Shafer (1976)] Shafer, G. *A mathemathical theory of evidence.* Princeton, Princeton University Press, 1976.

[Dempster (1967)] Dempster, A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping. *Annals Mathematics Statistics*, 38, 1967, p. 325-339.

[Dempster (1967a)] Dempster, A.P. Upper and Lower Probability Inferences Based on a Sample from a Finite Univariate Population. *Biometrika*, 54, 1967, p. 515-528.

[Duda et al. (1976)] Duda, R.O.; Hart, P.E. & Nilsson, N.J. Subective Bayesian methods for rule- based inference systems. *AFIPS Conference Proceedings*, N.Y., June 1976, p. 1075-1082.

[Zadeh (1978)] Zadeh, L. A. Fuzzy Sets as a Basic for a Theory of Possibility. *Fuzzy Sets and Systems 1*, 1978, p. 3-28.

[Zadeh (1965)] Zadeh, L. A. Fuzzy Sets. *Information and Control*, 1965, p. 8:338-353.

[Reiter (1980)] Reiter, R. A Logic for Default Reasoning. *Artificial Intelligence 13*, 1980, p. 81-132.

[Pearl (1982)] Pearl, J. Reverend Bayes on infeeeerence engines: a distributed hierachical approach. *Proceedings os the Second National Conference on Artificial Intelligence*, Pittsburgh, PA, 1982, p. 133-136.

[Pearl (1985)] Pearl, J. "How To Do With Probabilities What People Say You Can't". *IEEE Computer Society Press*, Los Alamitos, Calif., Dec. 1985, p. 6-12.

[Cheeseman (1985)] Cheeseman, P. In Defense of Probability. *Proc. Ninth IJCAI*, Palo Alto, Calif., 1985, p. 1002-1009.

[Pawlak (1982)] Pawlak, Z. Rough sets. *International Journal of Computer and Information Sciences*, 1982, 11(5):341-356.

[Lindley (1987)] Lindley, D.V. The Probability Approach to the Treatment of Uncertainty in AI and Expert Systems. *Statistical Science*, Vol.2, N.1, 1982, p.17-24.

[Wise & Henrion (1986)] Wise, B.P & Henrion, M. A Framework for Comparing Uncertain Inference Systems to Probability. *Uncertainty in AI*, New York, N.Y., 1986, p.69-83.

[Horvitz et al.(1986)] Horvitz, E.J & Heckerman $ Langlotz, C.P. A Framework for Comparing Alternative Formalisms of Plausible Reasoning. *Proc. Fifth Nat'l Conf. AI*, p.210-214, 1986.

[Gordon & Shortliffe (1984)] Gordon, J. & Shortliffe, E.H. The Dempster-Shafer Theory of Evidence. *Rule-based expert systems*, New York, Addison-Wesley, 1984, p. 272-292.

[Szolovits (1978)] Szolovits, P. & Pauker, S.G. Categorical and Probabilistic Reasoning in Medical Diagnosis. *AI, Vol.11, N.1-2*, 1978, p. 115-144.

[Heckerman (1986)] Heckerman, D. Probabilistic Interpretations For MYCIN's certainty factors. *Uncertainty in AI*, New York, L.N. Kanal and J.F. Lemmer eds., Elsevier Science Publishers, 1986, p. 167-196.

[Wang (1994)] Wang, P. A defect in Dempster-Shafer theory. *Technical Report N. 85 of CRCC*, Indiana, Indiana Universty, 1994.

[Stein (1993)] Stein, R. The Dempster-Shafer theory of evidential reasoning. *AI Expert*, (August): 26-31, 1993.

[Uchôa (1998)] Uchôa, J. Q. *Representação e indução de conhecimento usando teoria de conjuntos aproximados.* São Carlos, UFSCar, 1998. 237p. (Dissertação de Mestrado).

[Uchôa & Nicoletti (1997)] Uchôa, J. Q. & Nicoletti, M. C. *Elementos da Teoria de Dempster-Shafer*, Relatório Técnico do Departamento de Computação 07/97, São Carlos, DC-UFSCar, 1997, 34 p.

[Santos & Nicoletti] Santos, M. O. & Nicoletti, M. C. *O Uso do Modelo de Bayes em Sistemas Baseados em Conhecimento*, Relatório Técnico do Departamento de Computação, São Carlos, DC-UFSCar, 19 p.

[Santos & Nicoletti] Santos, M. O. & Nicoletti, M. C. *O Uso de Fatores de Certeza em Sistemas Baseados em Conhecimento*, Relatório Técnico do Departamento de Computação, São Carlos, DC-UFSCar,19 p.