
SOBRE O PROBLEMA DA INTEGRAÇÃO GENERALIZADA DE DADOS

Milton B. Do Coutto Filho*

mbrown@ic.uff.br

Julio C. Stacchini de Souza†

julio@ic.uff.br

Marcus Th. Schilling*

schilling@ic.uff.br

*Instituto de Computação - Universidade Federal Fluminense
Rua Passo da Pátria, 156 - Bloco E, sala 350 - CEP 24210-040 - Niterói, RJ - Brasil

†Depto de Eng. Elétrica - Universidade Federal Fluminense
Rua Passo da Pátria, 156 - Bloco E, sala 350 - CEP 24210-040 - Niterói, RJ - Brasil

ABSTRACT

This paper presents a set of methods which use quantitative and qualitative data, gathered from various heterogeneous sources of information, aiming at combining them in a process of data fusion. Different forms of uncertainty representation and modelling are discussed. Also, this work proposes a taxonomy for the data fusion processes according to the nature and redundancy of the processed data.

KEYWORDS: Data Fusion; Uncertainties; Fuzzy Systems; Intelligent Systems; Statistical Models; Belief Models.

RESUMO

Este artigo reúne um conjunto de métodos que utilizam dados de natureza quantitativa e qualitativa, provenientes de fontes heterogêneas, de modo a combiná-los em um processo conhecido por fusão de dados. Diversas formas de incerteza e sua modelagem são discutidas. Propõe-se uma taxionomia em quatro níveis para processos de fusão de dados, tomando-se por base a natureza e redundância dos dados processados.

Artigo submetido em 21/11/2004

1a. Revisão em 14/07/2006

2a. Revisão em 22/10/2006

3a. Revisão em 27/12/2006

Aceito sob recomendação do Editor Associado

Prof. Ivan Nunes Da Silva

PALAVRAS-CHAVE: Fusão de Dados; Incertezas; Sistemas Nebulosos; Sistemas Inteligentes; Modelos Estatísticos; Modelos de Credibilidade.

1 INTRODUÇÃO

Um mecanismo de integração de dados compõe-se de um conjunto de métodos que, utilizando dados provenientes de fontes de distintas naturezas, destina-se a combinar tais dados para melhorar a qualidade da informação deles extraída (Delen, 1998; Barbosa *et alii*, 2000).

O conceito de fusão de dados foi introduzido no início da década de 70 em pesquisas médicas destinadas à produção de diagnósticos via métodos computacionais, assim como no campo militar em estudos para identificação de alvos (aéreos, terrestres, aquáticos). Desde então, o problema da fusão de dados esteve relacionado à necessidade de se combinar dados precisos (natureza binária) e imprecisos (nebulosos), que apresentem diversos graus de incerteza (Antony, 1995).

O cérebro humano pode ser considerado como um dos mais eficientes dispositivos inteligentes de integração de dados. Vários sinais de natureza distinta - coletados pelos órgãos dos sentidos, contendo informações heterogêneas, especializadas, com certo grau de confiabilidade - são nele continuamente processados. Tais sinais são interpretados e coerentemente combinados, de modo a extrair dos dados imprecis-

tos disponíveis a informação considerada como mais crível (considerando-se os limites impostos pelo processo de medição e coleta de dados, materializados pelos órgãos dos sentidos), usualmente denominada realidade percebida.

A fusão de dados busca reconstruir, atualizar e avaliar informações oriundas de qualquer parte do mundo físico de interesse, através da combinação de informações provenientes de fontes de observação de cunho sensorial, factual e de conhecimento (Pan *et alii*, 1998). Assim, o modelo deste mundo torna-se o ponto central de um sistema de fusão de dados. Este modelo provê o espaço comum em que todos os objetos esparsos e as fontes de informação presentes são devidamente registradas, mapeadas e catalogadas de forma organizada. Os processos de combinação de dados são particularmente necessários quando múltiplos sinais e elementos sensores heterogêneos estão envolvidos. Desta forma, transições do tipo sinal-símbolo se fazem essenciais e o elemento informação detém o foco do processo de fusão de dados. Tais processos são construídos para monitorar e controlar sistemas físicos, interagindo com operadores humanos (Hall e McMullen, 2004).

Os métodos científicos comumente utilizados para a combinação de informações - provenientes de dados quantitativos (numéricos) - são baseados em algoritmos que atribuem ponderações para as informações disponíveis, e.g.: redes neurais artificiais, Filtro de Kalman e procedimentos estatísticos de natureza Bayesiana (Hall e Llinas, 2001).

Para situações em que se utilizam dados qualitativos (lingüísticos), cuja informação usualmente apresenta-se de modo incompleto, impreciso, contaminada por incertezas, os modelos clássicos de fusão de dados não são suficientes. Assim, recomenda-se o uso de metodologias especiais, representadas na Figura 1 com suas sobreposições (Klir e Yuan, 1995), tais como: Teoria da Evidência e da Possibilidade de Dempster-Shafer, Teoria da Alta/Baixa Probabilidade e Heurísticas Híbridas Inteligentes (e.g., modelos de credibilidade baseados em agentes e sistemas especialistas). Tais metodologias visam quantificar as incertezas presentes nos dados adquiridos em sistemas complexos, oriundas da confiabilidade parcial que se tem na informação, da imprecisão inerente à linguagem de representação na qual a informação é expressa, da não completude da informação e da agregação/sumarização da informação que provêm de múltiplas fontes. Incertezas de dois tipos são usualmente representadas: as de natureza aleatória, associadas à variabilidade dos dados de entrada e/ou parâmetros do modelo; e aquelas consideradas epistêmicas, subjetivas, relativas ao desconhecimento do processo/mecanismo que as governa. Exemplos de incertezas aleatórias são: variação de condições atmosféricas, variação da vida útil de equipamentos, etc. As incertezas epistêmicas estão presentes em situações em que, por

exemplo, exista falta de dados experimentais que caracterizem novos materiais e processos.

A escolha de um modelo apropriado para a quantificação de incertezas depende das características da informação disponível. Frequentemente, em diversas aplicações, os dois tipos de incerteza mencionados estão presentes. As incertezas aleatórias são adequadamente tratadas pela Teoria da Probabilidade. Já as incertezas epistêmicas podem ser tratadas com base na Teoria Nebulosa (funções de pertinência), Teoria da Evidência (funções de credibilidade e plausibilidade), Teoria da Possibilidade (medidas de possibilidade e necessidade), etc. Maiores detalhes sobre modelos para quantificação de incertezas serão apresentados ao longo do presente trabalho, em especial na Seção 5.

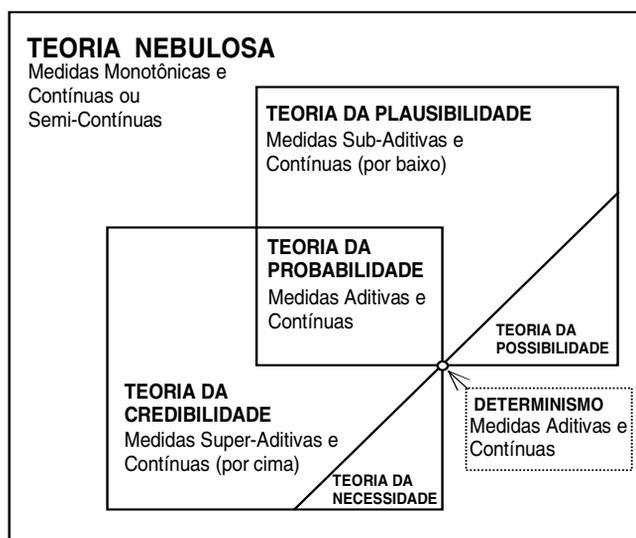


Figura 1: Teorias destinadas à quantificação das formas de incertezas presentes em diversos tipos de dados.

Os processos de fusão de dados numéricos com dados lingüísticos podem ser construídos por meio de ferramentas baseadas em heurísticas híbridas inteligentes (e.g., lógica neural-nebulosa), se a situação assim o permitir. Técnicas que avaliam diferentes formas de conhecimento parcial ou impreciso (Walley, 1997), inerentes à lógica neural-nebulosa (Klir e Yuan, 1995; Ross, 1995; Brown e Harris, 1994; Lin e Lee, 1996), vêm sendo desenvolvidas.

A adoção de uma abordagem unificada para o projeto de sistemas de decisão baseados na técnica de fusão de dados tem um impacto significativo no que diz respeito ao desempenho de tais sistemas. Delen (1998) identificou um conjunto de elementos-chave para o desenvolvimento de aplicações baseadas em processos inteligentes de fusão de dados. Segundo esta referência, todo processo que utilize procedimentos de fusão de dados deve conter:

- Representação padronizada dos dados a serem combinados;
- Mecanismo que traduza convenientemente os vários tipos de dados para a padronização adotada;
- Metodologias iterativas (com base na engenharia de *software* e *hardware*) para o desenvolvimento de sistemas inteligentes de fusão de dados;
- Formas de extração do conhecimento a partir dos dados qualitativos e quantitativos com base em modelos computacionais;
- Métodos que suportem múltiplos modelos de integração de dados;
- Mecanismos destinados à captura dos dados baseados em uma análise racional do sistema;
- Processos para a validação dos resultados obtidos.

Assim, um processo de fusão de dados usualmente se inicia com a atuação de um mecanismo de tradução que realize o mapeamento de dados brutos coletados de um sistema qualquer para o formato padronizado de representação escolhido (homogeneização). Em seguida, os dados traduzidos são interpretados e combinados por mecanismos inteligentes de fusão de dados (e.g. baseados em modelos que caracterizem os estados do sistema), normalmente implementados através de algoritmos de otimização que apresentam como objetivo a combinação dos dados (e.g. estimação de estado). Por fim, um mecanismo inteligente de diagnóstico atua, a fim de validar os resultados do processo de fusão de dados, fornecendo a informação de melhor qualidade possível (a mais provável e crível) sobre estado do sistema.

Este artigo reúne um conjunto de métodos que utilizam dados de natureza quantitativa e qualitativa, provenientes de fontes heterogêneas, de modo a combiná-los em processos conhecidos por fusão de dados. Tais métodos são encontrados de forma dispersa e não concatenada na literatura, dificultando a obtenção de uma idéia mais ampla e genérica sobre o tema. Assim, além de reunir de forma sistematizada os principais métodos de fusão, o artigo propõe uma taxionomia para os processos de fusão de dados, tomando por base o grau de complexidade do problema em estudo. Diversas formas de incerteza e sua representação são discutidas. As seções que se seguem tratam do problema da integração generalizada de dados de forma ampla, buscando fornecer elementos básicos para o ingresso nesta área de pesquisa. A Seção 2 apresenta uma taxionomia dos processos de fusão de dados. As Seções 3 e 4 tratam, respectivamente, do problema da fusão de dados quantitativos redundantes e complementares. A Seção 5 apresenta a fusão de dados apenas qualitativos e a Seção 6 da fusão de dados quantitativos e qualitativos. A Seção 7 aborda

perspectivas para o problema da integração generalizada de dados.

2 CLASSIFICAÇÃO DOS PROCESSOS DE FUSÃO DE DADOS

Devido à ausência de uma classificação genérica para os sistemas de fusão de dados visando aplicações, usualmente se adota uma classificação funcional de níveis hierárquicos, tomando-se por base as atividades relacionadas aos objetivos diretamente associados ao processo de combinação dos dados. Entretanto, aqui propõe-se uma classificação genérica para os modelos de fusão de dados, adotando-se a ordem crescente do nível de abstração (complexidade) encontrado nos dados submetidos ao processo de fusão. Assim, na Tabela 1 propõe-se tal classificação, partindo-se da integração de dados redundantes, de natureza essencialmente quantitativa (1^o nível), até a combinação de dados heterogêneos, de naturezas quantitativa e qualitativa (4^o nível).

Para ilustrar os tipos de dados que participam de um processo de fusão, considere por exemplo o problema da Estimação de Estado em Sistemas de Potência (Coutto Filho *et alii*, 2004). Realiza-se tal processo de estimação com o sistema operando em regime quase-permanente, sendo suficiente definir seu estado operativo como composto pelas tensões complexas de todos os nós da rede elétrica de interesse (magnitudes e ângulos de fase). Dados quantitativos redundantes, correspondentes a observações do estado, são coletados por unidades terminais remotas em toda a rede e transmitidos para um Centro de Operação de Sistemas para processamento. Usualmente, estes dados são medidas de grandezas elétricas, tais como: fluxos de potência nos ramos da rede, potências injetadas por fontes externas e magnitudes de tensão. Representa-se a incerteza de cada medida por um ruído Gaussiano de média zero e desvio conhecido, correspondente ao grau de precisão dos equipamentos necessários a sua aquisição. Para estimar o estado do sistema, processos distintos de integração de dados podem ser construídos, dependendo de diversas situações que a seguir são caracterizadas, ordenadas de acordo com o grau de complexidade que representam (conforme estabelecido na Tabela 1).

1^o Nível: Os dados quantitativos coletados são suficientemente redundantes para a estimação do estado da rede, realizada utilizando-se um método de otimização tal como o dos Mínimos Quadrados Ponderados.

2^o Nível: De modo a enriquecer o processo de estimação, dados quantitativos complementares aos coletados podem ser agregados, como no caso de se dispor de previsões de estado (estimativas *a priori*). Um processo de estimação com base no Filtro de Kalman pode ser construído.

Tabela 1: Classificação em níveis crescentes de complexidade para Sistemas de Fusão de Dados

Níveis	Características dos dados	Categorias das técnicas	Meios de implementação	Exemplos
1º	Integração de dados quantitativos redundantes	Métodos Estatísticos para combinação de informações redundantes - Princípio da Máxima Probabilidade.	<i>Hardware</i> dedicado	Dispositivos multisensores de natureza local micro-especializada, multiprocessada e tolerante a falhas.
2º	Integração de dados quantitativos de natureza complementar	Métodos Estatísticos para combinação de informações complementares – Teo. de Bayes; Princípio do Máximo <i>a Posteriori</i> ; Métodos Inteligentes e Híbridos baseados na extração e ponderação do conhecimento quantitativo.	<i>Hardware</i> microprocessado e tecnologia de <i>software</i>	Dispositivos multisensores dispersos; Redes Bayesianas; Redes Probabilísticas Nebuloso-Causais; Redes Nebulosas Bayesianas; Fusão Estática; Fusão Contextual; Redes Neurais; Lógica Nebulosa; Modelos de Estruturas para a Fusão de Dados; Estratégias de Votação; Lógica Neural-Nebulosa; Diagramas Piramidais; Filtro de Kalman; Raciocínio Indutivo; Segmentação Markoviana.
3º	Integração de dados qualitativos	Métodos Inteligentes e Híbridos baseados na extração e ponderação do conhecimento qualitativo - Regras de Dempster-Shafer; Eng. do Conhecimento; Funções de Crédito; Variáveis Nebulosas; Intervalos de Confiança.	<i>Hardware</i> multiprocessado e tecnologia de <i>software</i>	Sistemas Especialistas; Modelos de Multi-Agentes (Nebulosos) Inteligentes; Modelos de Credibilidade; Modelos de Nebulosidade; Estratégias de Votação; Modelos Híbridos de Raciocínio Indutivo.
4º	Integração de dados quantitativos e qualitativos	Metodologias Altamente Especializadas	<i>Hardware</i> multiprocessado e tecnologia de <i>software</i>	Lógica Neural-Nebulosa; Heurísticas Especializadas Híbridas Inteligentes.

3º Nível: Este seria um caso incomum para a estimação de estado em sistemas de potência, hoje uma tarefa extremamente importante em Centros de Operação, em que nenhum dado quantitativo estaria disponível. Os dados operativos do sistema disponíveis seriam caracterizados por perfis de comportamento histórico, experiência de operadores, faixas de valores de grandezas elétricas, etc. Sistemas especialistas, por exemplo, são capazes de agregar as informações disponíveis para produzir uma estimativa do estado operativo do sistema. Variáveis nebulosas podem representar as incertezas contidas nestas informações.

4º Nível: Neste caso deseja-se integrar aos dados quantitativos coletados outros de natureza qualitativa (e.g., aqueles citados no nível anterior), como por exemplo, em razão da ocorrência de indisponibilidade de unidades terminais remotas, perda de dados por problemas no

sistema de transmissão de informações, eliminação de dados espúrios, etc., ou simplesmente para enriquecer o processo de estimação com mais informações (Barboza, 1999, 2000). Os casos com maior potencial para a aplicação de fusão de dados de 4º nível seriam aqueles referentes a redes de distribuição de energia elétrica, onde torna-se extremamente oneroso instalar medidores com grau de redundância suficiente para se realizar a estimação de estado (Miranda *et alii*, 1995).

Deve-se mencionar que uma etapa importante do processo de estimação de estado é a validação dos dados processados. De modo a garantir que as estimativas obtidas para as grandezas de interesse para a supervisão do sistema não estejam corrompidas por dados de entrada espúrios (e.g. medidas portadoras de erros grosseiros, dados desatualizados da configuração da rede), usualmente testes com os resíduos da

estimação são realizados (Abur e Expósito, 2004).

3 FUSÃO DE DADOS QUANTITATIVOS REDUNDANTES

O desenvolvimento de sistemas de integração de dados de 1º nível vem apresentando um grande avanço na área de engenharia. Comumente constituídos por células de multisensores, tais sistemas destinam-se à coleta de medidas consideradas importantes para o conhecimento/controlar de sistemas complexos em tempo-real. O *hardware* associado a estes torna o processo de coleta tolerante a falhas, devido à alta redundância envolvida em projetos desta natureza. Nestes sistemas de integração de dados, atuando conjuntamente com o *hardware*, encontram-se algoritmos de fusão de dados implementados em nível local, que objetivam aumentar a qualidade da informação transportada pelo dado quantitativo coletado.

Considere que o *hardware* do sistema multisensor ilustrado na Figura 2 seja constituído por N sensores que medem de modo paralelo e síncrono uma dada grandeza. Associado a este *hardware* deve haver um algoritmo que promova a fusão dos dados provenientes dos sensores, considerando as características individuais dos referidos elementos de medição. O vetor que representa cada sensor inclui classificadores estatísticos que parametrizam a confiabilidade destes (classe de precisão, calibração, etc). Considere então um sistema estatístico MISO (*Multiple Input Single Output System*), sendo seu circuito combinatorial constituído por um mecanismo de fusão de dados que leve ao estabelecimento de uma regra de decisão. A modelagem matemática para o circuito combinatorial deste sistema pode ser desenvolvida a partir da aplicação do princípio estatístico da máxima probabilidade (Helene e Vanin, 1981).

Goodman e Michael (1996) apresentaram metodologias para a integração de sistemas de 1º nível compostos por processos multisensores que apresentam funções de caracterização gaussianas. Em Barbosa (1999) e Fischl e Kam (1991) encontram-se exemplos de aplicações em projetos na área de supervisão e controle de sistemas de energia elétrica.

4 FUSÃO DE DADOS QUANTITATIVOS COMPLEMENTARES

O uso de mecanismos integradores de dados de 2º nível vem se consolidando com diversas aplicações de cunho prático, devido à grande capacidade de monitorar, controlar e automatizar sistemas complexos de modo integrado (Waltz, 1990; Varshney e Burrus, 1997; Mutambara, 1998). Tal processo de fusão de dados concorre para o aumento da qualidade da informação pós-processada, obtida a partir da combi-

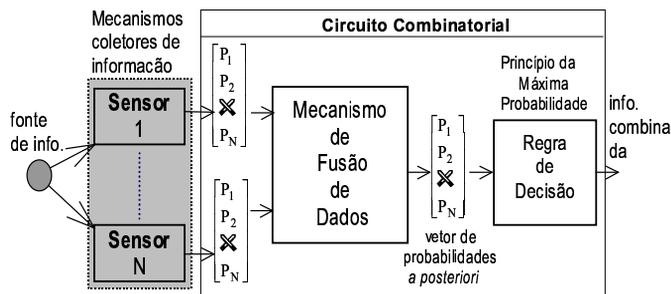


Figura 2: Modelo estatístico de combinação de dados quantitativos redundantes (fusão de dados de 1º nível)

nação de informações complementares advindas de diversos mecanismos de integração de dados de 1º nível, atuantes no sistema em estudo (Varshney e Burrus, 1997).

Na Figura 3, apresenta-se o esquema genérico de um sistema multisensor de aquisição e fusão de dados de 2º nível, constituído por M subsistemas de 1º nível que recebem, em nível local, de forma síncrona e paralela, determinado conjunto heterogêneo de medidas de caráter complementar. Associado a este *hardware* deverá haver um algoritmo de combinação de informações que promova a fusão dos dados provenientes dos subsistemas de multisensores, levando em conta as características individuais dos dados heterogêneos coletados, bem como as particularidades de cada um dos subsistemas de multisensores de fusão de 1º nível utilizados.

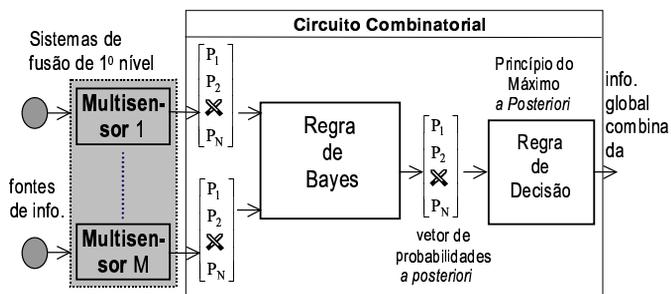


Figura 3: Modelo Bayesiano de combinação de dados quantitativos complementares (fusão de dados de 2º nível)

Processos estatísticos associados à aplicação do Teorema de Bayes (Gelman *et alii*, 1995) compõem o núcleo dos circuitos combinatoriais de integração e agregação dos dados. Em um contexto bayesiano, são realizadas combinações das informações provenientes dos múltiplos vetores de dados. O vetor que representa cada conjunto de informação combinada inclui classificadores estatísticos que medem a probabilidade da informação agregada como um vetor de parâmetros. Estes parâmetros são regidos por probabilidades condicionais. Assim, utiliza-se a Regra de Bayes para obter a composição da probabilidade *a posteriori*. Aplica-se então uma regra de

decisão, tal como o Princípio do Máximo *a Posteriori* (Hall e Mc Mullen, 2004; Goodman *et alii*, 1997), para selecionar a combinação de informações mais apropriada.

Pan *et alii* (1998) e Heckerman *et alii* (1994) exploraram os potenciais de aplicação das modernas formas de representação do conhecimento propiciado por redes bayesianas em mecanismos multisensores de fusão de dados de 2º nível. A seguir, a título de ilustração, a Figura 4 apresenta um esquema simplificado de uma destas aplicações.

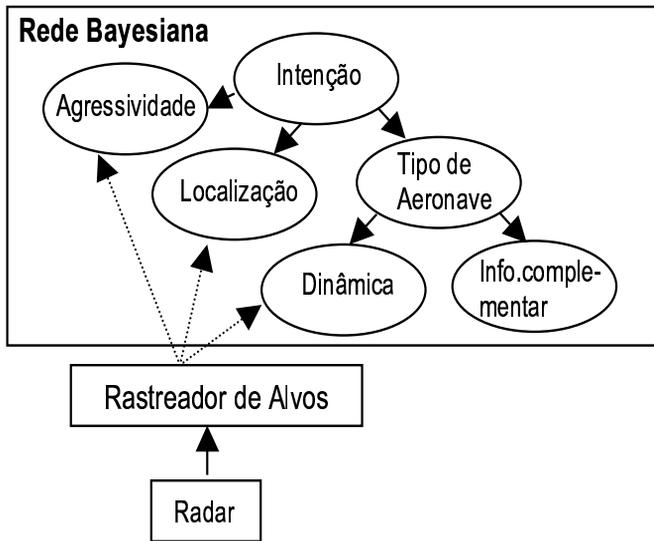


Figura 4: Modelo de uma rede bayesiana, destinada a obter o tipo do alvo rastreado e estimar suas intenções.

Na Figura 4 retrata-se uma aplicação militar, referente ao problema de vigilância aérea, em que uma rede bayesiana é usada para obter probabilidades para o tipo de alvo rastreado e suas intenções (Pan *et alii*, 1998). Os possíveis tipos de alvo são: aeronave de combate; helicóptero; míssil; jato de linhas aéreas; e aeronave civil de pequeno porte. As intenções de cada alvo estão restritas a: amigável; neutra; e hostil. Assume-se que, em um primeiro nível de processamento representado pelo rastreador de alvos, dados coletados por um sistema de radar (e.g., posição, velocidade, aceleração) com suas respectivas incertezas são integrados (vetor de parâmetros) e alimentam os nós da rede. Algumas regras aplicáveis ao problema, estabelecidas como probabilidades condicionais (P), são descritas a seguir:

- P (dinâmica da aeronave | tipo de aeronave), depende das restrições de velocidade e aceleração dos tipos de aeronaves, assim como das estimativas e respectivos erros da velocidade e aceleração do alvo;
- P (informação complementar | tipo de aeronave), considerando o uso de informação adicional (e.g., caracte-

rística Doppler de helicópteros) para determinar o tipo de aeronave;

- P (tipo de aeronave | intenção), com base no fato de que forças hostis usam aeronaves específicas para ataque.
- P (localização | intenção), usando informação de posicionamento para determinar a probabilidade da aeronave estar em uma zona militar restrita ou atravessando um corredor aéreo civil.

Barboza (1999) abordou a supervisão de redes de energia elétrica através de um processo de estimação de estado, com base no método dos mínimos quadrados ponderados (equações normais) para a fusão de dados provenientes de multisensores heterogêneos que apresentam funções de caracterização gaussianas. Em Coutto Filho *et alii* (2004), encontra-se uma abordagem para o problema de estimação de estado, em que se considerou um modelo de previsão para a dinâmica do sistema. Este foi construído a partir de dados históricos de operação, para obter estimativas *a priori* do comportamento de curtíssimo prazo que do estado do sistema. Tal algoritmo de fusão de 2º nível apresenta como principal benefício a possibilidade de validação *a priori* dos dados recebidos para processamento pelo estimador. Com isto, pode-se criar um processo de inovações, comparação estatística entre valores estimados *a priori* (previsões) e medições, que facilita a identificação de anormalidades nos dados a processar (notadamente, a presença de medidas portadoras de erros grosseiros; erros de configuração da rede sob supervisão; mudança brusca do estado operativo), bem como aperfeiçoa o processo de filtragem de incertezas. Em Souza *et alii* (1996), encontram-se algoritmos para o diagnóstico destas anormalidades e dos resultados do estimador. A Figura 5 ilustra um processo de fusão de dados destinado à estimação de estado em sistemas de potência.

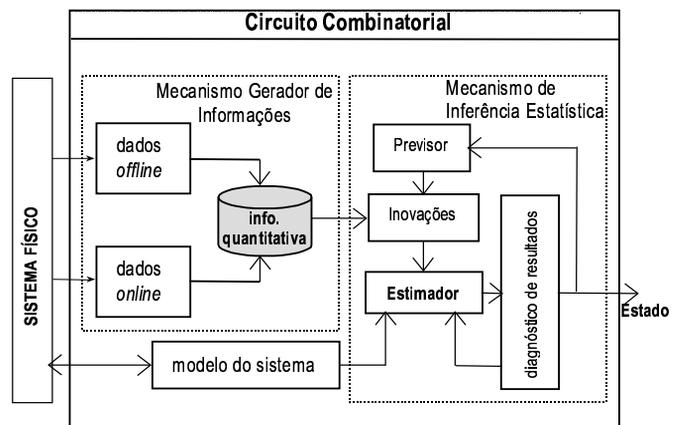


Figura 5: Sistema integrado de aquisição e fusão de dados destinado à supervisão de sistemas de potência.

Harris *et alii* (1997) exploraram os potenciais benefícios de aplicação de lógica neural-nebulosa para o desenvolvimento de modelos de fusão de dados de 2º nível utilizando mecanismos multisensores. Antony (1995) estudou a representação do conhecimento e sua manipulação por intermédio de métodos de raciocínio indutivo, para direcionar projetos de algoritmos destinados a manipulação de dados e a estruturação de bases de dados multimídia. Tais bases são úteis ao desenvolvimento de modelos de automação que utilizam tecnologias de fusão de dados de 2º nível. Gens *et alii* (1997) usaram técnicas de fusão de dados para o desenvolvimento de algoritmos voltados à combinação de imagens de alta e baixa definição. Prinz *et alii* (1997) propuseram a utilização de algoritmos de fusão de dados de 2º nível para a geração de imagens de alta definição, resolução e nitidez, a partir dos dados multiespectrais coletados por satélites orbitais.

5 FUSÃO DE DADOS QUALITATIVOS

Genericamente, dados de natureza qualitativa podem ser combinados com base no conhecimento extraído de fontes de informações subjetivas (e.g., consultas a especialistas, agentes informantes, sistemas multisensores especializados). Vários fatores interferem no processo de coleta da informação, tais como: grau de confiabilidade do canal de comunicação (biológico ou tecnológico), convicção, ponto de vista, crédito, formação religiosa e cultural, apegos pessoais do agente informante. Desta forma, o uso de ferramentas heurísticas inteligentes híbridas de apoio à decisão, destinadas a extração e combinação do conhecimento, proveniente de diversas fontes imprecisas de informação faz-se necessário (Bouchon-Meunier, 1997).

Na área de inteligência artificial duas técnicas vêm se consagrando como heurísticas destinadas à elaboração de modelos quantitativos para a extração/combinção do conhecimento de fontes qualitativas de informação imprecisa: (i) sistemas especialistas (Jackson, 1990; Gebhardt *et alii*, 1998; Giarratano e Riley, 2004); (ii) modelos de credibilidade (Smets, 1994; Kellogg e Gmytrasiewicz, 1997; Smets, 2000) baseados na modelagem de sistemas de multi-agentes inteligentes (Khosla e Dillon, 1997; Giunchiglia *et alii*, 1998; Jain *et alii*, 2002). Cada uma dessas técnicas tem seu domínio de aplicação bem definido, relacionado ao tipo da informação disponível e ao grau de explicitabilidade do mapeamento funcional que se deseja utilizar para extrair dos dados qualitativos disponíveis o conhecimento a eles agregado.

Os sistemas especialistas objetivam transladar os fenômenos observados, tal como se apresentam, para realidades concebidas, de modo a possibilitar o uso desta percepção para a tomada direta de ações e decisões. Tais sistemas são elaborados para coletar e agrupar fatos através do senso comum e da observação de dados empíricos, de modo a formular hipó-

teses, conduzir testes para a validação e/ou rejeição destas, com base em regras. Em outras palavras, sistemas especialistas podem ser considerados como parte de uma entidade maior que objetiva avaliar a situação baseando-se no exame de entradas de estímulo e na aprendizagem e elaboração de uma tese para a propagação das influências, tomando em consideração evidências adicionais para a elaboração e agrupamento de regras ou raciocínios lógicos (Giarratano e Riley, 2004).

Os modelos de credibilidade apresentam orientações de ordem implícita e explícita para os dados empíricos. As percepções apresentam suas regras de formação e combinação menos observáveis do que as dos sistemas especialistas. Em um modelo de credibilidade, convicções e orientações podem estar baseadas em condicionamentos, credos, aspectos culturais ou apegos pessoais de cunho subjetivo apresentados pelos agentes inteligentes; ou, até mesmo, em elementos sutis, de natureza pessoal, que globalmente sejam pouco relevantes, mas que localmente fazem-se importantes para um ou outro agente inteligente tomar decisões (Padgham e Wini-koff, 2004). Para os agentes, o processo ou ato de orientação pode ser influenciado por questões diversas, tais como: herança genética, expectativas sociais e tradições culturais dos agentes. De forma oposta, nos sistemas especialistas (de caráter mais científico) há menos influência de elementos culturais/irracionalis ou de fatores não comprováveis. Porém ambos, sistemas especialistas e modelos de credibilidade, são importantes para a construção de sistemas humanos de decisão (Brachman e Levesque, 2004).

5.1 Sistemas Especialistas e Agentes Inteligentes

Nesta seção serão apresentados conceitos básicos sobre sistemas especialistas e agentes inteligentes, de forma a dar elementos para uma melhor compreensão sobre os processos de integração de dados de natureza qualitativa.

5.1.1 Sistemas Especialistas

Um sistema especialista se caracteriza por: (a) lidar com problemas complexos que requerem a interpretação de um especialista humano; (b) solucionar estes problemas por meio de um modelo computacional que reproduza a conduta do raciocínio de um especialista; (c) atender a uma aplicação específica e limitada do conhecimento humano.

Os especialistas são profissionais competentes na solução de certos tipos de problemas. Usualmente, tal competência resulta de ampla experiência, adquirida na avaliação de soluções propostas para problemas corriqueiros de sua atividade profissional. Um sistema especialista computacional deve ser

dotado de suficiente conhecimento do especialista humano, de modo a também ser capaz de solucionar problemas específicos com perícia. Conhecer como o cérebro trabalha na solução de problemas especializados não é condição necessária para o projeto de sistemas especialistas. Basta ser possível extrair, agrupar e estruturar conhecimentos de especialistas, de modo a que estes possam ser processados computacionalmente.

Para avaliar um determinado problema, o especialista usa fatos, adota hipóteses e usa o conhecimento que possui, muitas vezes fruto de anos de vida profissional. Na Figura 6, apresentam-se componentes de um sistema especialista e observam-se os principais cursos de raciocínios utilizados.

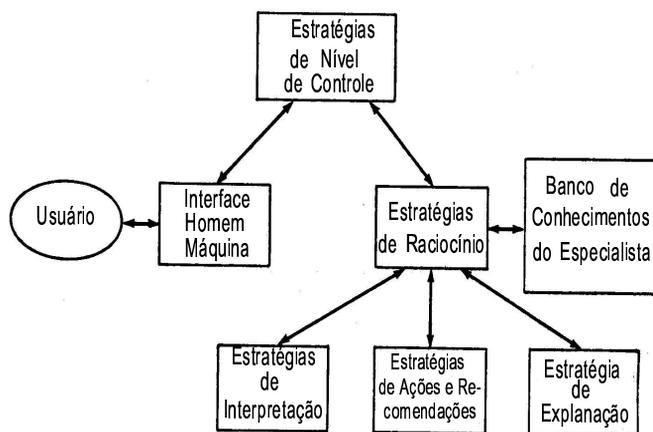


Figura 6: Principais cursos do raciocínio de um sistema especialista.

Durante o processo de raciocínio, o especialista avalia a importância dos fatos associados a determinado problema, por comparação com informações referentes ao conhecimento prévio acumulado em situações semelhantes. Neste processo, vão-se formulando novas hipóteses e verificando-se fatos; e esses irão influenciar o processo de raciocínio. Em alguns casos, o especialista pode não chegar a um diagnóstico conclusivo, se os fatos de que dispõe para aplicar seu conhecimento anterior não forem suficientes. Um sistema especialista deve, além de inferir, ter capacidade de aprender a partir da aquisição de novos conhecimentos e, desse modo, melhorar seu desempenho.

De um modo geral, os sistemas especialistas produzem três formas de diagnóstico: (a) delimitam o universo de possíveis respostas para o problema em estudo; (b) determinam um pequeno número de respostas possíveis; (c) apenas interagem com quem os consulta. Sempre que um problema não puder ser algoritmizado ou sua solução conduzir a um processamento complexo e demorado, tais sistemas podem ser uma boa opção, dada a sua intrínseca natureza heurística.

Um sistema especialista pode ter o conhecimento armazenado em si advindo da experiência de diversos especialistas de determinada área. Este conhecimento poderá ser preservado, agrupado e transmitido a outros profissionais desta área, se constituindo assim em poderosa ferramenta de treinamento.

Um sistema especialista processa conhecimentos e informações; não processa dados. O conhecimento é reunido e armazenado em uma base de própria e os dados são ajustados contra ela. O processamento é feito adotando-se este conhecimento acumulado. Em uma análise mais ampla, um sistema especialista compõe-se de três elementos básicos: (a) quadro-negro; (b) mecanismo de inferência; (c) base de conhecimento.

O quadro-negro, também conhecido como rascunho, é uma área da memória onde o sistema grava e apaga os dados usados no processo de inferência, até chegar a um diagnóstico para o problema formulado. Assim sendo, há necessidade de se avaliar regras que são recuperadas da base de conhecimento para uma área de trabalho. Neste local, tais regras são ordenadas periodicamente para avaliação, verificando-se fatos e hipóteses. Também há necessidade de uma área onde possam ser guardados os valores dessas variáveis para se trabalhar tais fatos e hipóteses. As conclusões dessas regras irão gerar novos fatos e novas hipóteses que precisam ser guardados temporariamente durante o processo de inferência ou dedução. A área destinada à execução dessas operações chama-se quadro-negro ou rascunho.

A base de conhecimento norteia o funcionamento do sistema especialista, i.e., tendo sido projetada para receber informações de uma determinada área do conhecimento, o sistema será especialista nessa área. A base de conhecimento destina-se a armazenar fatos e regras. Um novo fato pode modificar todo um processo de inferência de acordo com as regras existentes sobre ele que estão sendo aplicadas e também sobre os novos fatos gerados a partir da avaliação dessas regras. Redes probabilísticas bayesianas (Pearl, 1988), comumente utilizadas para a representação do conhecimento em problemas tais como os de diagnóstico médico (Diéz *et alii*, 1997), apresentam um modelo gráfico com relações causais que buscam aproximar os mecanismos utilizados naturalmente por especialistas em sua prática cotidiana.

O conhecimento do especialista é armazenado de forma padronizada, definida pelo projetista do sistema, de tal modo que este conhecimento possa ser tratado de uma maneira sistemática. Este conhecimento normalmente é passado para o sistema através de um editor, construído especialmente para ser usado pelo especialista na transferência do seu conhecimento para o sistema. O editor também formata o conhecimento extraído de uma maneira apropriada, conforme a

representação projetada para o sistema armazenar esse conhecimento. Se o profissional especialista colocou regras de boa qualidade, o desempenho do sistema deverá refletir essa qualidade fazendo inferências com alto grau de acerto e de maneira muito rápida, dando boas justificativas para suas respostas. Uma base de conhecimento é sempre feita para uma área bem definida do conhecimento.

O mecanismo de inferência é o elemento de um sistema especialista capaz de buscar as regras necessárias a serem avaliadas, ordenadas de uma maneira lógica e, a partir daí, ir direcionando o processo heurístico de inferência. O processo de encaminhar a inferência é feito de acordo com a técnica que foi usada para o armazenamento do conhecimento na base de conhecimento. O método mais utilizado é o da avaliação de regras.

Um sistema especialista ideal deve ser concebido considerando-se os seguintes módulos:

- a) **Processador de Linguagem Natural:** Estabelece uma interface homem-máquina que se assemelhe à linguagem natural, evitando que o usuário do sistema necessite aprender uma linguagem de programação específica para uso do mesmo. Na realidade, este módulo é composto por um subsistema especialista.
- b) **Justificador do Conhecimento:** Interage com o usuário indicando o caminho que levou o sistema a determinada conclusão ou o porquê do sistema estar fazendo uma determinada pergunta. Utiliza recursos e estruturas próprias, exibindo regras e fatos da base de conhecimento, sempre que for solicitado.
- c) **Quadro-Negro:** Compõe a área de trabalho do sistema, i.e., escreve, apaga, tira conclusões que levem a um diagnóstico final. Compreende três áreas lógicas:
 - **Planejamento:** figuram as regras da base de conhecimento a serem avaliadas;
 - **Agenda:** ordenam-se as regras a serem avaliadas;
 - **Solução:** armazena-se a solução obtida durante o processo de inferência e situam-se as estruturas de controle necessárias à inferência.
- d) **Base de Conhecimento:** Ordena e armazena (com base em regras semânticas ou lógica de predicados) adequadamente fatos e regras a serem utilizados pelo mecanismo de inferência e pelo módulo justificador de conhecimento. Deseja-se que a base de conhecimento tenha a capacidade de aprender, a partir de treinamentos com especialistas, ou através de textos e experiências.
- e) **Seqüenciador:** Escolhe as regras que serão selecionadas a partir da reunião dos fatos e hipóteses existentes, bem

como determina a ordem em que essas regras serão avaliadas.

- f) **Interpretador:** Avalia as regras em função das hipóteses/fatos existentes, o que gera novos fatos e hipóteses que por sua vez selecionarão novas regras até se chegar a uma conclusão final.
- g) **Reforçador da Consistência:** Verifica a solução final alcançada pelo sistema, observando se a mesma está consistente com outras regras existentes na base de conhecimento, assim como se existe algum caso similar resolvido. Cataloga as soluções/conclusões dos novos casos para uso futuro.

5.1.2 Agentes Inteligentes

Define-se agente inteligente como um componente de *software* e/ou de *hardware* que apresenta a capacidade de agir, reagir e interagir, de um modo coerente, e em conformidade com as mudanças que ocorram no ambiente externo próximo ao agente. Tais mudanças podem estar relacionadas a ações de outros elementos (de natureza humana ou não) presentes nas proximidades do meio que cerca o agente. Na realidade, o termo agente é genérico, englobando todo um conjunto de tipos específicos de agentes. Uma tipologia funcional para a classificação dos agentes pode ser convenientemente elaborada a partir das características exibidas pelos agentes e utilizada para diferenciar diversos grupos dessas entidades. Assim, os agentes podem ser classificados de acordo com sua mobilidade (estáticos e móveis), i. e., sua habilidade em se movimentar ao longo ou em torno de uma determinada rede. Também podem ser classificados com relação a sua capacidade decisória (deliberativos e reativos). Agentes deliberativos apresentam internamente um modelo heurístico, simbólico, de raciocínio artificial, de tal modo que esse grupo de agentes pode interagir entre si e/ou com o meio ambiente que o cerca de um modo planejado, negociado e coordenado. Opostamente, os agentes reativos não são dotados de quaisquer modelos de raciocínio artificial, atuando unicamente por meio de procedimentos do tipo estímulo-resposta com outros agentes e/ou com o meio ambiente que os cerca.

Sete categorias básicas de agentes podem ser catalogadas, segundo as funções exercidas pelos mesmos: (1) colaborativos; (2) de interface; (3) móveis; (4) de informação; (5) reativos; (6) híbridos; (7) inteligentes. Estas categorias, embora sejam utilizadas para descrever o desenvolvimento e a intenção dos agentes, não são necessariamente independentes; podem haver agentes que apresentem sobreposição funcional de atividades. Certos atributos primários também podem ser utilizados para a classificação de agentes, tais como, sua autonomia e capacidade de cooperação. Autonomia é o princípio pelo qual um agente pode operar sozinho sem a necessidade de

um direcionamento (humano); o componente chave da autonomia é o nível de pró-atividade, ou a habilidade de tomar iniciativa, independentemente da pré-existência de estímulos externos ao agente (e.g., aqueles provenientes do ambiente). O princípio de cooperação é de suma importância. Na disposição de cooperar, os agentes necessitam exibir uma habilidade social, i.e., uma pré-disposição para interagir entre si e, possivelmente, com humanos. A manifestação da característica de habilidade social têm como pré-requisito o estabelecimento de alguma forma de comunicação entre as partes. Para isso, deve haver um canal de comunicação e uma linguagem comum entre as partes envolvidas. Porém, mesmo sem dispor de procedimentos de cooperação complexos, formalmente estabelecidos, os agentes podem coordenar suas ações simplesmente respondendo a estímulos ambientais. Para que os sistemas baseados em agentes sejam verdadeiramente classificados como inteligentes, eles devem ser capazes de aprender com a experiência e reagir/interagir com o ambiente externo.

Agentes também podem ser classificados através de sua função de gerenciamento e organização, tais como, os de informação universal (*WorldWide Information Agents*). Estes agentes ajudam a administrar a vasta quantidade de informação que trafega pela rede de comunicação que interliga os agentes. Existem muitos outros atributos e comportamentos relevantes que devem ser considerados em um agente. Assim, por exemplo, um agente pode ser considerado como versátil (possui muitas metas ou apresenta diversidade de comportamentos); pode ser benevolente ou maleficiente; antagonista ou altruísta; pode apresentar um comportamento maligno ou sincero; pode ser um agente confiável (ou de confiança, no sentido de delegar uma dada tarefa a esse agente) ou não-confiável; pode apresentar um comportamento rítmico (no sentido de realizar uma dada atividade de um modo contínuo) ou arrítmico; pode apresentar um comportamento entusiástico ou depressivo (degrada sistematicamente suas atividades e enfraquece perante certas dificuldades); podem ser preguiçosos ou ativos, etc. Atributos emocionais também podem ser atribuídos aos agentes. Eles podem ficar cansados, zangados, irritados, saudosos (do lar ou de outros agentes), etc. Alguns agentes também podem exibir comportamentos associados a atitudes e conceitos de natureza mental, tais como, um certo grau de credibilidade, de desejo e de intenção.

Finalmente, os agentes são classificados como híbridos quando apresentam uma combinação de características ou acumulam tarefas de agentes de duas ou mais classes. Esta diversificação de características implica em que o conceito de agente somente possa existir em um contexto multidimensional, onde estão envolvidos altos níveis de abstração.

Outra questão a ser observada diz respeito ao fato de que um

agente não apresenta, por definição, um caráter de benevolência para com outro agente. Um comportamento benevolente entre dois ou mais agentes (ou grupos de agentes) surge através da competição entre os mesmos e/ou entre os agrupamentos destes. Tal comportamento deriva da necessidade de realização de alguma atividade, sob a forma de cooperação, com a finalidade de obter algum ganho, de natureza local, que induza a uma diferenciação do grupo. Neste contexto são relegados momentaneamente para níveis secundários os comportamentos antagônicos entre os agentes e/ou entre agrupamentos de agentes que fazem parte do grupo. Pesquisas nesta área indicam que a classe dos agentes fortemente competitivos é normalmente formada a partir de representantes de todas as outras classes e tipos de agentes presentes no sistema.

Os sistemas de agentes heterogêneos são aqueles que apresentam em sua composição dois ou mais tipos de agentes pertencentes a diferentes classes. Um sistema heterogêneo de agentes pode conter em seu escopo um ou mais agentes híbridos. O universo de um sistema de agentes heterogêneos é provido por uma rica diversidade de produtos de *software* e *hardware*, os quais provêm uma grande diversidade de serviços geograficamente distribuídos em regiões do universo que compõem o sistema. Embora esses produtos trabalhem de forma local/regionalmente isolada, há sempre uma demanda incremental para operarem entre si, visto que um comportamento coletivo tende a alcançar melhores resultados do que a soma de comportamentos individuais, dada a superaditividade dos eventos de natureza possibilística. O requisito-chave para a interoperação entre agentes heterogêneos está centrado na linguagem de comunicação existente entre os mesmos, que deve apresentar protocolos de comunicação e mecanismos de negociação mediante os quais diferentes tipos de agentes podem se comunicar e interagir uns com os outros.

Procedimentos de engenharia de *software* baseados em modelos de agentes caracterizam um método moderno e radicalmente diferente para o projeto, implementação e manutenção de sistemas de computação flexíveis e inteligentes de tempo-real denominado sistemas orientados-a-agentes (Figura 7). O desenvolvimento desta tecnologia proporciona uma integração multidisciplinar entre os ramos dos sistemas de informação, da engenharia de *software*, da tecnologia de banco de dados multimídia e da inteligência artificial, aproximando-se dos chamados sistemas de inteligência artificial distribuída.

Os trabalhos de pesquisa sobre sistemas de agentes inteligentes heterogêneos requerem o uso e o desenvolvimento de metodologias, ferramentas, técnicas, padrões de gerenciamento e manipulação de dados capazes de interoperar em ambientes amorfos, permeados por fontes de informação heterogêneas, com distintos graus de confiabilidade.

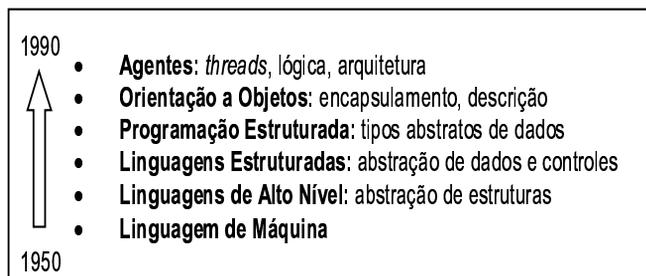


Figura 7: Posicionamento dos agentes no contexto da programação de computadores.

Os sistemas especialistas são adequados para a solução de problemas cujo domínio é bem delimitado e conhecido, tendo capacidade limitada de extrapolar o conhecimento nele codificado. Por sua vez, os agentes inteligentes são capazes de realizar inferências em um ambiente dinâmico, onde adaptações e interações com o meio que os cercam são necessárias. Muitas aplicações práticas requerem o processamento de informações de diferentes naturezas e a execução de procedimentos distintos que se complementam para a obtenção de uma solução. Certas partes de um problema podem ser resolvidas por procedimentos rotineiros e completamente conhecidos, situação em que o emprego de um sistema especialista se torna adequado. Em outros casos, quando decisões em um ambiente dinâmico são requeridas, ferramentas com capacidade de adaptação, tais como os agentes inteligentes, devem ser empregadas. Desta forma, a solução mais adequada para uma grande quantidade de problemas pode requerer a atuação conjunta das técnicas aqui mencionadas.

5.2 Modelos de Credibilidade

Os modelos de credibilidade (*Transferable Belief Models* - TBM) destinam-se à modelagem de problemas onde o atual estado de credibilidade não é exatamente conhecido, porém o são todos os estados possíveis de opinião que um dado agente do sistema pode ocupar (Smets e Kennes, 1994). Isto permite que as transições de estado de um agente sejam expressas de um modo axiomático e quantificadas através de uma teoria bem fundamentada da durabilidade, da credibilidade e da imprecisão intrínseca vinculada à opinião adotada por um dado agente, em um dado momento do tempo, comumente chamada de crédito ou credibilidade da informação. Interações entre múltiplos agentes - sejam eles de natureza interna ou externa ao sistema, além do fato de nunca conhecerem toda a informação disponível - elevam a complexidade do modelo TBM, explicitando não-linearidades, características de sistemas complexos. Os modelos TBM são mais gerais e flexíveis do que os modelos probabilísticos, visto que podem representar estados do conhecimento e de opiniões que os modelos probabilísticos não podem (Smets e Kennes, 1994).

A principal característica das funções de credibilidade (Shafer, 1990) reside no fato de que a união de dois eventos de crédito disjuntos pode resultar em um crédito maior ou igual do que a soma dos dois créditos, quando estes são tomados individualmente. Saffiotti *et alii* (1998) assinalam que todo modelo TBM é dotado dos seguintes elementos:

- Regra de transição de opiniões: cientificamente estabelecida e axiomatizada que viabiliza a implementação de um mecanismo de transição de estados opinativos, acarretada pelo conhecimento de novos fatos relevantes por parte do agente em questão;
- Regra de combinação de opiniões: uma nova opinião é induzida em um agente quando este toma conhecimento de duas (ou mais) evidências distintas;
- Regra de redução de ponderação de opiniões: permite uma reavaliação das informações recebidas de fontes com certo grau de confiabilidade;
- Regra de inversão de funções de crédito condicionais: utilizada como uma extensão do Teorema de Bayes;
- Regra para a geração de medidas de probabilidade: necessária para computar uma medida de utilidade presumida que viabiliza a determinação de decisões otimizadas;
- Estratégias para a implementação de processos de inferência estatística: e.g., análise de *cluster*, análise de discriminantes, mecanismos de diagnose, etc.

Os modelos TBM podem ser utilizados em diversos domínios de aplicação, tais como: (i) fusão de dados qualitativos gerados por fontes de informação com certo grau de confiabilidade; (ii) fusão de opiniões de especialistas; (iii) métodos de classificação e análise de discriminantes; (iv) análise de *clusters*; (v) diagnose; (vi) geração de bases de dados contendo dados imprecisos, qualitativos, armazenando informações disjuntas; (vii) construção de mapeamentos; (viii) atividades de raciocínio indutivo.

Embora à primeira vista os modelos TBM não aparentem diferir dos modelos probabilísticos, na realidade são dotados de capacidades não disponíveis pelos modelos baseados na teoria da probabilidade clássica. Isto justifica o interesse crescente da comunidade científica em desenvolver e sistematizar tais modelos. Assim, os seguintes aspectos dos modelos de credibilidade devem ser ressaltados:

- a) Capacidade para representar qualquer estado de parcial ou total ignorância;

- b) Trabalha com a informação momentaneamente disponível, não necessitando da produção de probabilidades para aqueles eventos que não se encontram disponíveis para o uso. As ponderações são introduzidas para os eventos imediatamente disponíveis para uso;
- c) Não há necessidade de se conhecer qualquer probabilidade *a priori* em processos de diagnóstico. O conhecimento *a priori* pode ser representado por todas as formas de conhecimento relativamente ao estado de total ignorância;
- d) O sistema pode realizar diagnósticos mesmo enquanto apresenta categorias indefinidas, e.g. uma classe que retrata uma desordem de natureza desconhecida ou a existência de algum estado que não foi catalogado;
- e) Na análise de discriminantes, pode-se utilizar no conjunto de treinamento dados para os quais a categoria de pertinência é imprecisamente conhecida;
- f) Nas bases de dados, o sistema pode manusear tanto dados errados como também dados que apresentam valores imprecisos;
- g) Nas atividades de aproximação do raciocínio o sistema pode manusear tanto o conhecimento “comum”, considerado confiável, quanto o conhecimento “ponderado”.

As incertezas associadas à informação conduzida por um agente podem ser modeladas utilizando-se as regras explicitadas na lógica nebulosa. A lógica nebulosa apresenta a capacidade de manipular, de modo quantitativo, diferentes tipos de incertezas (e.g., vaguidão, nebulosidade, obscuridade, indistinção, dissonância, incongruência, discrepância, conflito, imprecisão, credibilidade, ambigüidade, não-especificidade, variedade, generalidade, diversidade, equívoco e discórdia). Variáveis lingüísticas são utilizadas para associar valores a funções de caracterização de modo a prover níveis métricos para os vários tipos de incertezas presentes. Na pesquisa avançada em teoria da informação a combinação dos modelos de agentes inteligentes com a capacidade de representação das técnicas de modelagem de incertezas provenientes da lógica nebulosa culmina na elaboração de modelos computacionais de agentes inteligentes nebulosos, elementos de pesquisa recente que determinam um desenvolvimento dos mais promissores (Bell, 1998).

Diversas formas de representação podem ser utilizadas para quantificar incertezas associadas à informação processada, a saber: probabilidades imprecisas (Smithson, 1989, 1997; Walley, 1997) e comparativas ordenadas (Fine, 1973); intervalos de confiança, também chamados de probabilidades altas, baixas ou não-aditivas (Good, 1962); funções de credibilidade (Dempster, 1967; Shafer, 1976, 1990); avaliação de

capacidades (Huber, 1981; Denneberg, 1994); variáveis nebulosas; medidas de possibilidades; conjuntos de medidas de probabilidades (Levi, 1980); altas e baixas previsões (Williams, 1976; Walley, 1996); modelos classificatórios (Fine, 1977; Walley e Fine, 1979); conjuntos de risco (Williams, 1976; Walley, 1991); ordenamento parcial de preferências (Giles, 1976; Giron e Rios, 1980).

Correntemente, o Teorema de Bayes e a Teoria de Dempster-Shafer, da Evidência e da Possibilidade (Dempster, 1967; Shafer, 1976; Zadeh, 1978; Dubois e Prade, 1988, 1993; Guan e Bell, 1992; Klir e Harmanec 1994; Yager *et alii*, 1994; Klir e Yuan, 1995; Kohla e Monney, 1995) compõem o núcleo teórico utilizado para a descrição e simulação matemática dos métodos inteligentes de suporte à decisão. Enquanto probabilidades (no sentido clássico) estão associadas a eventos que representam chances, probabilidades imprecisas caracterizam a informação de natureza qualitativa (imprecisa).

A seguir, são apresentadas duas estruturas básicas de modelos de credibilidade para a fusão de dados de 3º nível.

5.2.1 Modelo de Dempster-Shafer

Este modelo baseia-se nas premissas matemáticas desenvolvidas pela Teoria da Evidência e da Possibilidade de Dempster-Shafer, onde cada hipótese pode ser representada por dois conjuntos de parâmetros:

- a) **Índice de Sustentabilidade:** $0 \leq S(x) \leq 1$; também denominado medida de credibilidade ou da evidência; representa o grau que um conjunto de medidas sustenta ou evidencia uma dada hipótese.
- b) **Índice de Plausibilidade:** $0 \leq P(x) \leq 1$; também denominado medida de possibilidade ou da necessidade; representa o grau em que a evidência falha em rejeitar a hipótese. Definindo-se $S(\bar{x})$ como o índice de sustentabilidade do complemento de x , obtém-se $P(x)$ para um dado elemento x pelo complemento de $S(\bar{x})$, ou seja: $P(x) = 1 - S(\bar{x})$. A diferença $I(x) = P(x) - S(x) = 1 - [S(x) + S(\bar{x})]$ é uma medida da ignorância acerca da hipótese. Quando $P(x) = S(x)$, o intervalo de probabilidades colapsa para uma medida de probabilidade simples equivalente a probabilidade condicional *a posteriori*.

Na Figura 8 apresenta-se o diagrama esquemático do modelo de fusão de dados de 3º nível de Dempster-Shafer. Nele considera-se um sistema constituído de M classificadores e/ou sensores especializados que coletam N dados qualitativos de natureza complementar, a partir de fontes de informação com certo grau de confiabilidade. O vetor que representa cada sensor/classificador contém os índices de sus-

tentabilidade e plausibilidade da informação coletada. Estes parâmetros quantificam a probabilidade imprecisa dos dados fornecidos pelos sensores/classificadores e representam, em certo sentido, medidas de hipóteses. Em um contexto qualitativo, esses parâmetros são regidos por evidências e possibilidades. Assim, uma regra de combinação, análoga à Regra de Bayes, provê o meio de se computar a composição de intervalos de sustentabilidade e plausibilidade (ou intervalos de credibilidade) de cada hipótese, reduzindo as incertezas implícitas dos dados medidos. Uma regra de decisão é então aplicada tendo como base os índices de sustentabilidade/plausibilidade de modo a selecionar a combinação de informações mais apropriada.

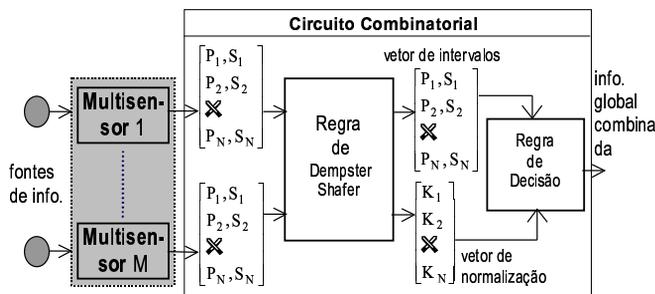


Figura 8: Modelo de Dempster-Shafer para a combinação de informações qualitativas complementares (fusão de dados de 3º nível).

5.2.2 Modelo Estruturado de Fusão

Recentemente, foi proposta uma metodologia para a fusão de dados de 3º nível (Waltz e Buede, 1986) que busca modelar a informação a partir de um processamento de fluxos e incorporar a teoria dos conjuntos nebulosos como modo de medição das incertezas. Esta metodologia divide o processamento de incertezas em três estágios, com se descreve a seguir.

- **Estágio 1:** Um conjunto de M classificadores/sensores especializados coletam N dados qualitativos, de natureza complementar, a partir de fontes de informação com certo grau de confiabilidade. Neste estágio, assume-se que o modelo opera em um sistema de tempo discreto onde os classificadores/sensores provêm como saída um vetor de parâmetros. Este é composto por elementos individuais de informação S_i e por um valor de função de caracterização u_i , computado neste estágio para cada elemento individual de informação, de modo a caracterizar a qualidade da informação imprecisa adquirida através da avaliação de sua credibilidade individual (u_i).
- **Estágio 2:** Aqui são computadas, a partir da teoria dos

conjuntos nebulosos, três medidas de incerteza (armazenadas em vetores de parâmetros) que caracterizam diferentes formas de imprecisão da informação coletada no estágio 1. A Figura 9 representa esquematicamente as referidas incertezas. Valores das funções de caracterização são processados por um mecanismo nebuloso de avaliação de regras que opera em duas etapas, de modo a analisar, avaliar, agregar e fundir os valores das funções de caracterização dos elementos individuais de informação coletados.

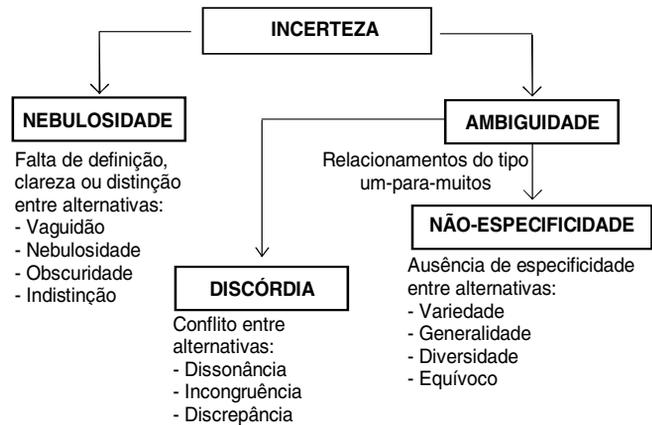


Figura 9: Três tipos básicos de incertezas na informação

Na 1ª etapa, um conjunto de regras nebulosas de seleção é utilizado, com a finalidade de avaliar se a informação coletada por um determinado sensor apresenta valores em sua função de caracterização compatíveis com a situação em questão. Isto é realizado por meio de cortes de nível- α nas funções de caracterização dos dados coletados, onde os valores de caracterização assim obtidos são transformados em intervalos. Uma função de avaliação coerentemente adotada para a situação identifica se os intervalos são apropriados aos casos que representam. Na 2ª etapa, determinam-se os valores de três tipos de incertezas associadas a cada um dos elementos de informação.

A saída do 2º estágio compõe-se de um conjunto de três vetores de parâmetros que quantificam as seguintes incertezas:

- a) **Não-Especificidade ou Imprecisão:** $V_i = (U_1, U_2, \dots, U_n)$. Esta grandeza resulta da necessidade de quantificação de certos aspectos da incerteza, tais como: variedade, generalidade, diversidade, equívoco/imprecisão. Ela busca quantificar o grau de representatividade da informação corrente quando comparada a outros possíveis conjuntos de informações relevantes, por intermédio de uma medição vinculada ao tamanho (ou cardinalidade) do conjunto amostral.

A medição desta componente da incerteza torna-se necessária visto que provê um modo de avaliar a representatividade da informação de um dado conjunto amostral com relação à informação total que está sendo fornecida pelos classificadores/sensores. A função U , definida através de uma função de Hartley (Klir e Yuan, 1995), provê um modo aproximado de avaliação da incerteza associada à adoção de conjuntos amostrais alternativos. Adotando a teoria de conjuntos clássicos, Hartley mostrou que a medida de não-especificidade está vinculada a uma classe de funcionais da forma: $U(A) = c \cdot \log_b |A|$, onde $|A|$ denota a cardinalidade do conjunto finito, não-vazio A , sendo b e c constantes, tais que, $b > 1$ e $c > 0$. Cada par de valores de (b,c) determina uma unidade em que a incerteza é medida. Para $(b,c) = (2,1)$, a incerteza é medida em *bits* e a função de incerteza assume a forma: $U(A) = \log_2 |A|$.

A generalização do funcional de Hartley, passando-se do domínio da Teoria dos Conjuntos convencional para a Teoria dos Conjuntos Nebulosos, foi proposta no início dos anos 80 sob a denominação de incerteza- U . Para qualquer conjunto nebuloso não-vazio A , definido sobre um conjunto universo finito X , a função de Hartley generalizada assume a forma:

$$U(A) = \frac{1}{h(A)} \int_0^{h(A)} \log_2 |\alpha A| d\alpha \quad (1)$$

onde $|\alpha A|$ denota a cardinalidade dos cortes de nível- α do conjunto A e $h(A)$ é a amplitude do conjunto (i. e.: o maior valor assumido pela função de caracterização que representa o conjunto). Observe que a não-especificidade de A , $U(A)$, é uma média ponderada de valores da função de Hartley para todos os distintos cortes de nível- α normalizados de $A(x)/h(A)$, $\forall x \in X$.

b) Discórdia: $W_i = (D_1, D_2, \dots, D_n)$. Esta grandeza resulta da necessidade de quantificação de uma componente da incerteza que caracterize a idéia de conflito (discórdia, incongruências, discrepâncias) entre duas ou mais alternativas relevantes. A parcela de incerteza associada à discórdia pode ser representada pelo maior valor da função de caracterização obtido usando-se operações de (re)união entre funções de caracterização que representem os conjuntos nebulosos.

c) Nebulosidade: $X_i = (F_1, F_2, \dots, F_n)$. Resulta da necessidade de quantificação da parcela de incerteza vinculada à noção de falta de definição, clareza ou distinção entre alternativas como fruto de aspectos de vaguidão, nebulosidade, obscuridade ou indistinção. Em geral, a medida da nebulosidade é obtida por meio da função $f: \mathfrak{S}(X) \rightarrow \mathbb{R}^+$, onde $\mathfrak{S}(X)$ denota o conjunto potência de

X . Para cada conjunto nebuloso A , esta função assume valores que expressam o grau de quanto à fronteira de A não está bem definida. Para que f represente bem a nebulosidade de um conjunto, três condições básicas devem ser satisfeitas:

- $f(A) = 0 \Leftrightarrow A$ for um conjunto convencional (*crisp*);
- $f(A)$ apresenta valor máximo em $A(x) = 0,5$, $\forall x \in X$ (de acordo com o mais alto grau de nebulosidade esperado);
- $f(A) \leq f(B)$, quando o conjunto A é indubitavelmente mais bem definido do que o conjunto B (significa que $A(x) \leq B(x)$, quando $B(x) \leq 0,5$ e $A(x) \geq B(x)$, quando $B(x) \geq 0,5$, $\forall x \in X$).

Demonstra-se que o grau de nebulosidade pode ser obtido a partir da utilização da distância de Hamming. Para se medir o grau de nebulosidade de um conjunto vale observar a distinção entre o conjunto e seu complemento. Quanto menos um conjunto diferir de seu complemento, mais nebuloso ele será. Na avaliação de Hamming, a distinção entre um dado conjunto A e seu complemento é dada por $|A(x) - [1 - A(x)]| = |2A(x) - 1|$, assim a ausência local de distinção será $1 - |2A(x) - 1|$. A medida da nebulosidade $f(A)$ pode ser obtida pela adição de todas as medidas locais de ausência de distinção:

$$f(A) = \sum_{x \in X} \{1 - |2A(x) - 1|\} \quad (2)$$

A Tabela 2 sumariza as características das categorias de incerteza computadas no processamento deste estágio.

Tabela 2: Incertezas em modelos de fusão de dados

Representação	Formas de avaliação	Tipos
Conjuntos Nebulosos	Funcional de Hartley	Não-Especificidade
Conjuntos Nebulosos	Operador de União	Discórdia e Conflito
Conjuntos Nebulosos	Distância de Hamming	Nebulosidade

- **Estágio 3:** Destina-se à avaliação da incerteza total presente na informação recebida, através da agregação das três formas de incerteza calculadas no estágio anterior. A medida da incerteza total de um sistema deve ser calculada a partir do uso do princípio da incerteza máxima.

Segundo este princípio, em qualquer processo de inferência a incerteza tem uma natureza essencialmente ampliada, e deve ser representada por operações do tipo máximo, supremo e união (de conjuntos nebulosos), de modo a se obter como resultado o mais alto nível de incerteza acumulada possível. No modelo estruturado de fusão de dados a incerteza total é representada por uma equação que assume a seguinte forma:

$$Q_i = \cup V_i. \cup W_i. \cup X_i \quad (3)$$

onde o operador \cup apresenta um contexto variável de tal forma que o princípio da incerteza máxima seja respeitado. Uma regra de decisão é então aplicada, tendo como base o índice de incerteza total de modo a selecionar a combinação de informações mais apropriada. Na Figura 10 apresenta-se o diagrama esquemático do Modelo Estruturado de combinação de informações qualitativas complementares, destinado ao desenvolvimento de sistemas de fusão de dados de 3º nível.

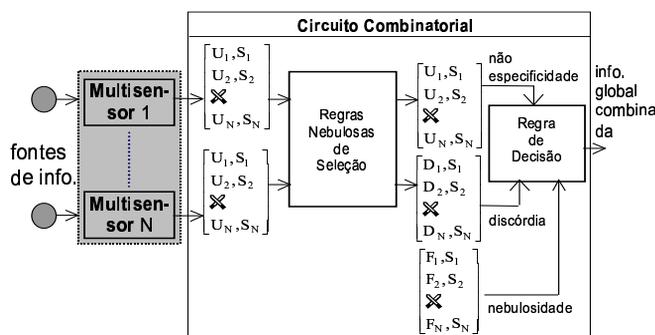


Figura 10: Modelo Estruturado de combinação de informações qualitativas complementares (fusão de dados de 3º nível).

6 FUSÃO DE DADOS QUANTITATIVOS E QUALITATIVOS

Processos de integração de dados numéricos (informação de natureza quantitativa) com dados lingüísticos (informação de natureza qualitativa) compõem o que aqui se denomina como fusão de dados de 4º nível. Atuando como sistemas integradores de mais alto nível podem apresentar como elementos de entrada dados provenientes de sistemas de fusão de 1º, 2º e 3º níveis.

Na Figura 11 apresenta-se um diagrama esquemático da arquitetura de um sistema de integração de informações de 4º nível. Um mecanismo integrador de informações heterogêneas recebe como entrada, de modo assíncrono, dados provenientes de uma base de dados multimídia composta por:

- dados quantitativos redundantes, oriundos de mecanismos de integração de informações de 1º nível;
- dados quantitativos complementares, oriundos de mecanismos de integração de informações de 2º nível;
- dados qualitativos, oriundos de mecanismos de integração de informações de 3º nível.

Neste contexto, o mecanismo integrador de informações heterogêneas promove um processo agregação e combinação das informações de entrada, de um modo cientificamente aceitável e sob algum aspecto otimizado, de modo a gerar como saída um conjunto de informações que apresente maior credibilidade, confiabilidade e nível de incerteza minimizado.

Mecanismos dedicados de fusão de dados de 4º nível vêm sendo amplamente utilizados em áreas da pesquisa científica de vanguarda para o desenvolvimento de aplicações no âmbito civil e militar, devido à grande potencialidade que apresentam (Heckerman *et alii*, 1994). Duas áreas merecem destaque:

- Aplicações que visam o aumento de qualidade/confiabilidade/credibilidade da informação final pós-processada, através da fusão de um conjunto heterogêneo de informações de natureza complementar, proveniente de diversos mecanismos de coleta e integração de dados de nível inferior que atuam ao longo do sistema em análise;
- Aplicações de natureza exploratória, que estejam intimamente voltadas para a descoberta de relações empíricas que vinculem efeitos provenientes da combinação de variáveis qualitativas com aqueles provenientes de (ou observados em) variáveis quantitativas, em sistemas muito complexos.

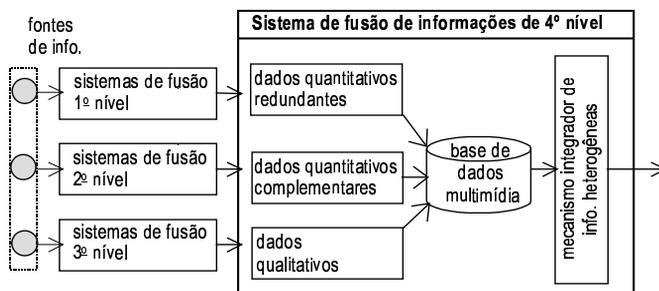


Figura 11: Diagrama esquemático de um sistema genérico de fusão de dados de 4º nível

Os modelos nebulosos permitem levar em consideração informações qualitativas fornecidas por especialistas humanos.

Por outro lado, os modelos de redes neurais, que apresentam dificuldades para representar informações qualitativas, possuem capacidade de generalização e tolerância a falhas, características altamente desejáveis para a execução de um mapeamento complexo, particularmente quando o domínio do problema não pode ser completamente coberto e representado pela perícia humana. Na Figura 12 observa-se um processo de combinação de informações heterogêneas, de natureza qualitativa e quantitativa, em que foram aplicadas técnicas inteligentes de extração do conhecimento, com base em métodos de raciocínio indutivo, sistematizados no âmbito da lógica neural-nebulosa.

Hart e Hoerger (1988) pesquisaram índices para avaliação do risco de desenvolvimento de carcinomas de diversos tipos com base na busca de relações empíricas mediante a realização de cruzamentos entre informações qualitativas – de origem psico-biológicas – e quantitativas – de origem biomédica, bioquímica, ambiental, genética, etc – de pacientes e de variáveis que caracterizam o meio-ambiente onde estes viveram ou vivem.

Lowenstein e Brick (1995) estudaram o desenvolvimento e a implantação de índices para avaliação da qualidade de vida de populações baseando-se no conceito de gerontologia ecológica, onde buscam encontrar relações entre os dados qualitativos e quantitativos vinculados aos aspectos sócio-econômicos, ambientais, tecnológicos, psicológicos e materiais tendo em vista a qualidade de vida das populações.

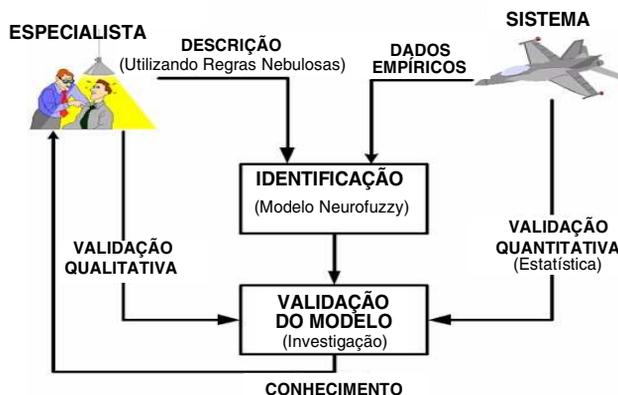


Figura 12: Fusão de dados numéricos (quantitativos) e lingüísticos (qualitativos) através da lógica neural-nebulosa.

Kanoum *et alii* (1997) pesquisaram a elaboração de metodologias combinadas que levem em consideração aspectos qualitativos (características da interação homem-máquina, qualidades das interfaces, etc) e quantitativos (metodologias de validação, de testes, grau de modularidade do projeto, etc) para o desenvolvimento de modelos de determinação da confiabilidade de *softwares*.

Gabbay *et alii* (1997) apresentaram uma coletânea de artigos que descrevem inúmeras técnicas interdisciplinares tais como: fundamentos de lógica monotônica, fundamentos de lógica modal, teoria das funções de credibilidade, redes Bayesianas, lógica nebulosa, teoria da possibilidade, algoritmos de inferência, etc, que são utilizadas para a elaboração de modelos estáticos e dinâmicos de combinação de dados qualitativos e quantitativos através do uso de técnicas extração do conhecimento mediante o uso de processos de raciocínio indutivo e de lógica neural-nebulosa.

Barboza (1999) pesquisou a elaboração de metodologias inteligentes e híbridas para a integração de informações de natureza qualitativa e quantitativa aplicada ao problema de Estimção do Estado de Sistemas de Potência de modo a determinar um estado para o sistema que apresente, simultaneamente, níveis de probabilidade e de credibilidade maximais.

Souza *et alii* (2004) apresentaram uma abordagem neural-nebulosa para o processamento de alarmes e identificação de componentes defeituosos em sistemas de potência. Relações *fuzzy* entre padrões de alarme e possíveis componentes defeituosos no sistema foram estabelecidas e utilizadas como conjuntos de treinamento para redes neurais. As redes neurais foram treinadas para produzir estimativas dos graus de pertinência dos componentes do sistema na classe de componentes defeituosos, tendo sido evidenciada a capacidade de se combinar o conhecimento qualitativo de especialistas, por meio da lógica fuzzy, com a capacidade das redes neurais de generalizar a partir do conhecimento adquirido.

7 PERSPECTIVAS

Neste artigo foram apresentados modelos destinados à tarefa de extração, representação e modelagem do conhecimento e da informação, de modo a tornar possível a integração de dados de naturezas distintas, através do uso de processos conhecidos por fusão de dados.

Para a combinação de informações, de forma diversa do ambiente probabilístico clássico, dados são representados por números imprecisos, tratados pela lógica nebulosa de forma flexível, no sentido de traduzir metas ou apresentar uma diversidade de comportamentos. Em um problema fusão de dados por combinação de informações nenhuma informação pode ser considerada tão ruim de modo que deva ser totalmente descartada do problema, e nenhuma informação pode ser considerada tão boa que deva ser totalmente considerada como essencial para o problema.

Em várias áreas do conhecimento humano tem sido caracterizado o conceito de sistema complexo (Amaral e Ottino, 2004; Bar-Yam, 2005), por vezes referenciado como: natureza fortemente não-linear; grande porte estrutural, não

necessariamente significando espaço físico; relações causa-efeito não completamente conhecidas; presença de variáveis discretas, contínuas, probabilísticas, possibilísticas, intervalares, etc. Exemplos clássicos de sistemas complexos são o corpo-humano, sistemas do meio ambiente, do clima, da economia de países, de energia elétrica.

Sabe-se também que a controlabilidade e observabilidade desses sistemas não se constituem em tarefas simples, pois para estas concorrem variáveis numéricas, difusas, linguísticas, visuais, sociais e informações heurísticas diversas. Assim sendo, torna-se claro que sistemas complexos são candidatos naturais a se valerem dos benefícios da aplicação das técnicas de fusão generalizada de dados, cuja proposta taxionômica foi pioneiramente abordada neste artigo.

Dentre os exemplos acima citados, sistemas de energia elétrica de grande porte constituem campos de pesquisa adequados à aplicação direta das técnicas de fusão de dados dessemelhantes. Isso se verifica porque existe inegável diversidade de variáveis e informações requeridas para a gestão adequada, segura e econômica desses sistemas (Mori, 2006; Ilic *et alii*, 2006).

Sabe-se, por exemplo, que a manutenção em tempo real da operação segura e econômica das grandes redes elétricas nacionais, envolve a manipulação de milhares de dados, amiúde conflitantes, imprecisos e de naturezas distintas. Além da característica eminentemente não linear, o espectro temporal dos fenômenos tratados é muito abrangente e as variáveis influentes têm naturezas discretas, contínuas, aleatórias, econômicas, regulatórias, sócio-políticas, ambientais, etc. Trata-se, portanto, de uma área de excelência para a aplicação direta da tecnologia de fusão generalizada de dados, tal como proposto no presente artigo.

Os problemas práticos em diversas áreas do conhecimento apresentam características que tornam apropriado o emprego de diferentes técnicas inteligentes, como aquelas aqui citadas. Pode-se afirmar que, para grande parte destes problemas, melhores soluções serão obtidas com o emprego de modelos híbridos, ou seja, aqueles que combinam mais de uma técnica de computação inteligente. Isto é particularmente interessante quando a aplicação de uma técnica isoladamente não é suficiente para tratar todos os aspectos de um dado problema. Dessa maneira, acredita-se que a utilização de modelos híbridos deverá nortear as aplicações de sistemas inteligentes aos diversos problemas do mundo real.

8 CONCLUSÕES

Este artigo apresentou um conjunto de métodos de manipulação de dados de natureza quantitativa (precisos) e qualitativa (imprecisos ou nebulosos) provenientes de fontes heterogê-

neas de informação, com o objetivo de agregá-los através de um processo de fusão de dados por combinação de informações de modo a melhorar a qualidade, a confiabilidade e a credibilidade da informação final assim obtida.

Os métodos apresentados ao longo do texto propõem uma representação da informação vinculada aos vários tipos de incertezas presentes em um problema, comumente desprezadas por não se saber como tratá-las matematicamente, de um modo natural e com um certo sentido lógico. Tais métodos tornam possível a integração de dados de diferentes naturezas de um modo coerente e otimizado.

A literatura técnica apresenta tais métodos de forma dispersa e não concatenada, dificultando aos pesquisadores interessados em fusão de dados a obtenção de uma idéia mais ampla e genérica sobre o problema. Assim, além de se reunir de forma sistematizada os principais métodos de fusão, propôs-se uma taxionomia em quatro níveis, que leva em consideração a natureza e redundância dos dados processados.

9 AGRADECIMENTOS

Os autores expressam seus agradecimentos ao CNPq e FAPERJ pelo apoio financeiro recebido e também a Mário César Schössler Barboza, cuja Dissertação de Mestrado forneceu o material para este artigo.

REFERÊNCIAS

- Abur, A. & A.G. Expósito (2004). *Power System State Estimation*, Marcel Dekker, New York.
- Agarwal, A.R., M.R. Tanniru & M. Dacruz (1992). Knowledge-Based Support for Combining Qualitative and Quantitative Judgements in Resource Allocation Decisions, *Journal of Management Information Systems*, Vol. 9, pp. 165-184.
- Amaral, L.A.N. & J.M. Ottino (2004). Complex Networks: Augmenting the Framework for the Study of Complex Systems, *European Physical Journal B*, Vol. 38, pp. 147-162.
- Antony, R.T. (1995). *Principles of Data Fusion Automation*. Artech House Publ., Boston.
- Barboza, M.C.S. (1999). Mecanismos para a Integração de Informações: Fusão de Dados Quantitativos e Qualitativos com Aplicação à Estimativa de Estado de Sistemas de Potência, *Dissertação de Mestrado*, Universidade Federal Fluminense, Pós-Graduação em Computação, Niterói, RJ.
- Barboza, M.C.S., M.B. Do Coutto Filho, M.Th. Schilling & J.C.S. Souza (2000). Fusão de dados heterogêneos,

- Anais do XIII Congresso Brasileiro de Automática*, Florianópolis, SC, pp. 796-801.
- Bar-Yam, Y. (2005). *Making Things Work: Solving Complex Problems in a Complex World*, Knowledge Press.
- Bell, P.R. (1998). The Development and Evaluation of Fuzzy Connections in C4ISR for Military Applications., *Defense Information Systems Agency*, Final Report, USA.
- Bouchon-Meunier, B. (1997). *Aggregation and Fusion of Imperfect Information*, Springer-Verlag.
- Brachman, R. & H. Levesque (2004). *Knowledge Representation and Reasoning*, The Morgan Kaufmann Series in Artificial Intelligence, Elsevier, San Francisco, CA.
- Brown, M. & C. HARRIS (1994). *Neurofuzzy Adaptive Modelling and Control*, Prentice Hall.
- Cavedon, L., A. Rao & W. Wobcke (1997). *Intelligent Agent Systems: Theoretical and Practical Issues*, Springer Verlag.
- Coutto Filho, M.B., J.C.S. Souza & M.Th. Schilling (2004). Previsões no Processo de Estimacão de Estado em Sistemas de Potência, *Anais do XV Congresso Brasileiro de Automática*, artigo 447, Gramado, RS.
- Delen, D. (1998). Integrated Data Experimentation & Fusion System, *Ballistic Missile Defense Organization*, USAF. <http://www.kbsi.com>.
- Dempster, A.P. (1967). Upper and Lower Probabilities Induced by a Multivalued Mapping. *Annals Mathematical Statistics*, Vol. 38, pp. 325-339.
- Denneberg, D. (1994). *Non-Additive Measure and Integral*, Kluwer, Dordrecht, 1994.
- Diéz, F. J., J. Mira, E. Iturralde & Z. Zubillage (1997). DIAVAL: A Bayesian Expert System for Echocardiography, *Artificial Intelligence in Medicine*, Vol. 10, pp. 59-73.
- Dongyong, Y. & Y. Yuzo (2000). Multi-sensor data fusion and its application to industrial control, *Proc. 39th SICE (Society of Instrument and Control Engineers) Annual Conference*, Iizuka, Japan, pp. 215-220.
- Dubois, D. & H. Prade (1988). *Possibility Theory*, Plenum Press, New York.
- Dubois, D. & H. Prade (1993). Fuzzy Sets and Probability: Misunderstanding, Bridges and Gaps, *Proc. 2nd IEEE Conf. Fuzzy Systems*, San Francisco, pp. 1059-1068.
- Fine, T.L. (1973). *Theories of Probability*. Academic Press, New York.
- Fine, T.L. (1977). An Argument for Comparative Probability. In: R. E. Butts, J. Hintikka (eds.) & D. Reidel, *Basic Problems in Methodology and Linguistics*, Dordrecht, pp. 105-119.
- Fischl, R. & M. Kam (1991). Advanced Measurement Communication and Information Systems Using Data Fusion Techniques. *Workshop on Real-Time Control and Operation of Electric Power Systems*, Denver, CO, pp. 43-54.
- Gabbay, D.R., Kruse, A. Nonnengart & H.J. Ohlbach, eds. (1997). Qualitative and Quantitative Practical Reasoning, *First Int. Joint Conf. on Qualitative and Quantitative Practical Reasoning*, ECSQARU-FAPR'97, Bad Honnef, Germany, Springer-Verlag, New York.
- Gebhardt, J., R. Kruse & C. Borgelt (1998). *Uncertainty and Vagueness in Knowledge Based Systems*. University of Magdeburg, Germany.
- Gelman, A., J.B. Carlin, H.S. Stern & D.B. Rubin (1995). *Bayesian Data Analysis*, Chapman & Hall.
- Gens, R., Z. Vekerdy & C. Pohl (1997). Image and Data Fusion: Concept and Implementation of a Multimedia Tutorial, *Proc. 2nd Int. Conf. on Fusion of Earth Data*, Sophia Antipolis, France, pp. 217-222.
- Giarratano, J.C. & G.D. Riley (2004). *Expert Systems: Principles and Programming*, 4th edition, Thomson.
- Giles, R. (1976). A Logic of Subjective Belief. In: W. L. HARPER, C. A. HOOKER (eds.) & D. REIDEL. *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Vol. 1, pp. 41-70, Dordrecht.
- Giron, F.J. & S. Rios (1980). Quasi-Bayesian Behaviour: A More Realistic Approach to Decision Making? (with discussion). In: J.M. Bernardo, M.H. DeGroot, D.V. Lindley & A.F.M. Smith (eds.). *Bayesian Statistics I*, Valencia University Press, Valencia, pp. 17-38, 49-66.
- Giunchiglia, F., N.R. Jennings & M. Wooldridge (1998). Special Issue on Computational and Logical Aspects of Multi-Agent Systems, *Journal of Logic and Computation*, Department of Computing, Imperial College of Science, Technology and Medicine, London.
- Good, I.J. (1962). Subjective Probability as the Measure of Non-Measurable Set. In: E. NAGEL, P. SUPPES & A. TARSKI (eds.), *Logic, Methodology and Philosophy of Science*, Stanford University Press, Stanford, pp. 319-329.

- Goodman, G.L. & D.W. McMichael (1996). A Gaussian Mixture Model Classifier Using Supervised and Unsupervised Learning, *4th Int. Symp. on Signal Processing and Its Applications*, Gold Coast, Australia, pp. 565-566.
- Goodman, I.R., R.P.S. Mahler & H.T. Nguyen (1997). *Mathematics of Data Fusion*, Kluwer Academic Publ.
- Guan, J.W. & D.A. Bell (1992). *Evidence Theory and Its Applications*, Vol. 2, Noth-Holland, New York.
- Hall, D.L. & J. Llinas (2001). *Handbook of Multisensor Data Fusion*, CRC Press LLC, Boca Raton, FL.
- Hall, D.L. & S.A.H. Mc Mullen (2004). *Mathematical Techniques in Multisensor Data Fusion*, 2nd edition, Artech House, Norwood, MA.
- Harris, C.J., Z.Q. Wu, K.M. Bossley & M. Brown (1997). Intelligent Neurofuzzy Estimators and Multisensor Data Fusion. In: S. G. Tzafestas (ed.), *Methods and Applications of Intelligent Control*. Netherlands, Kluwer Academic Publ.
- Hart, W. & F.D. Hoerger (1988). Carcinogen Risk Assessment: New Directions in the Qualitative and Quantitative Aspects, *Banbury Report 31*, National Center for Toxicological Research, The Dow Chemical Company, 1988.
- Heckerman, D., D. Geiger & D. Chickering (1994). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data, Technical Report MSR-TR-94-09, *Microsoft Research*.
- Helene, O.A.M. & V.R. Vanin (1981). *Tratamento Estatístico de Dados em Física Experimental*, Edgard Blucher, São Paulo.
- Huber, P. J. (1981). *Robust Statistics*, John Wiley & Sons, Inc., New York.
- Jain, L. C. , Z. Chen & N. Ichalkaranje (2002). *Intelligent Agents and Their Applications: Studies in Fuzziness and Soft Computing*, Physica-Verlag.
- Jackson, P. (1990). *Introduction to Expert Systems*. 2th Edition, Addison-Wesley Inc., New York.
- Kanoun, K., M. Kaaniche & J.C. Laprie (1997). Qualitative and Quantitative Reliability Assessment, *IEEE Software*, Vol. 14, No. 2, pp. 77-87.
- Kellogg, T. & P.J. Gmytrasiewicz (1997). Bayesian Belief Update in Multi-Agent Systems, *Workshop Machine Learning for User Modeling*, 6th International Conf. on User Modeling, <http://www.dfki.uni-sb.de/~bauer/umws/Final-Versions/Kellogg/node1.html>.
- Khosla, R. & T. Dillon (1997). *Engineering Intelligent Hybrid Multi-Agent Systems*, Kluwer Academic Pub.
- Klir, G. J. & D. Harmanec (1994). On Modal Logic Interpretation of Possibility Theory, *Int. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 2, No. 2, pp. 237-245.
- Klir, G.J. & B. Yuan (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall.
- Kohlas, J. & P.A. Monney (1995). *A Mathematical Theory of Hints: An Approach to the Dempster-Shafer Theory of Evidence*, Springer-Verlag.
- Levi, I. (1980). *The Enterprise of Knowledge*, MIT Press, London.
- Lin, C.T. & C.S.G. Lee (1996). *Neural Fuzzy Systems: A Neural-Fuzzy Synergism to Intelligent Systems*, Prentice Hall.
- Lowenstein, A. & Y. Brick (1995). The Differential and Congruent Roles of Qualitative and Quantitative Methods to Evaluate Quality of Life in Residential Settings, *Clinical and Experimental Research (Aging)*, Vol. 7, No.3, pp. 255-257.
- Ilic, M., J.W. Black, M. Prica (2006). Distributed Electric Power Systems of the Future: Institutional and Technological Drives for Near-Optimal Performance, *Electrical Power Systems Research*, in press, doi:10.1016/j.eprsr2006.08.013.
- Miranda, V., J. Pereira & J. Tomé Saraiva (1995). Experiences in State Estimation Models for Distribution Systems Including Fuzzy Data, *Proc. Stockholm Power Tech*, Stockholm, Sweden, pp. 288-293.
- Mori, H. (2006). State-of-the-Art Overview on Data Mining in Power Systems, paper 06GM0416, *IEEE Power Eng. Society General Meeting*, Montreal.
- Mutambara, A.G.O. (1998). *Decentralized Estimation and Control for Multisensor Systems*, CRC Press.
- Padgham, L. & M. Winikoff (2004). *Developing Intelligent Agent Systems: A Practical Guide*, Wiley Series in Agent Technology, Wiley.
- Pan, H., N. Okello, D. McMichel & M. Roughan (1998). Fuzzy Causal Probabilistic Networks and Multisensor Data Fusion, *Int. Symp. on Multispectral Image Processing (SPIE)*, Vol. 3543, Wuhan, China, pp. 1-12.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann.

- Prinz, B., R. Wiemker & H. Spitzer (1997). Simulation of High Resolution Sattelite Imagery from Multispectral Airborne Scanner Imagery for Accuracy Assentment of Fusion Algorithms, *Joint Workshop of ISPRS WG I/1, I/3 and IV/4, Sensors and Mapping from Space*, Hannover, Vol. 17, p.223-231.
- Ross, T. J. (1995). *Fuzzy Logic With Engineering Applications*, Mc.Graw-Hill, New York.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*, Princeton University Press.
- Shafer, G. (1990). Perspectives on the Theory and Practice of Belief Functions. *Int. J. Approx. Reasoning*, Vol. 4, pp. 323-362, 1990.
- Smets, Ph. & R. Kennes (1994). The Transferable Belief Model, *Artificial Intelligence*, Vol. 66, pp. 191-234.
- Smets, P. (2000). Data Fusion in the Transferable Belief Model, Proceedings of the 3rd International Conference on Information Fusion, Vol. 1, Paris, pp. 21-33.
- Smithson, M. J. (1989). *Ignorance and Uncertainty: Emerging Paradigms*, Springer-Verlag, Berlin.
- Smithson, M.J. (1997). Human Judgement and Imprecise Probabilities, *The Society for Imprecise Probability: Theories and Applications (SIPTA)*, pp. 1-10, www.sipta.org/documentation.
- Souza, J.C.S., E.B.M. Meza, M.B. Do Coutto Filho & M.Th. Schilling (2004). Alarm Processing in Electrical Power Systems Through a Neuro-Fuzzy Approach. *IEEE Transactions on Power Delivery*, Vol. 19, N° 2, pp. 537-544.
- Souza, J.C.S., A.M. Leite da Silva & A.P. Alves da Silva (1996). Data Debugging For Real-Time Power System Monitoring Based on a Pattern Analysis Approach, *IEEE Transactions on Power Systems*, Vol. 11, No. 3, pp. 1592-1599.
- Varshney, P.K. & C.S. Burrus (1997). *Distributed Detection and Data Fusion*, Springer Verlag.
- Walley, P. & T. L. Fine (1979). Varieties of Modal and Comparative Probability, *Synthese*, Vol. 41, pp. 321-374.
- Walley, P. (1991). *Statistical Reasoning With Imprecise Probabilities*, Chapman and Hall Press, London.
- Walley, P. (1996). Measures of Uncertainty in Expert Systems, *Artificial Intelligence*, Elsevier Science, Vol. 83, No. 1, pp. 1-58.
- Walley, P. (1997). *Imprecise Probabilities Project*, <http://ippserv.rug.ac.be>.
- Waltz, E.L. & D.M. Buede (1986). Data Fusion and Decision Support for Command and Control, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-16, No. 6, pp. 865 - 879 .
- Waltz, E. L. & J. Linas (1990). *Multisensor Data Fusion*, Artech House, Boston.
- Williams, P. M. (1976). Indeterminate Probabilities. In: M. Przelecki, K. Szaniawski, R. Wojcicki & D. Reidel (eds), *Formal Methods in Methodology of Empirical Sciences*, pp. 229-246, Dordrecht.
- Yager R.R., J. Kacprzyk & M. Fedrizzi (1994). *Advances in the Dempster-Shafer Theory of Evidence*, John Wiley & Sons, New York.
- Zadeh, L. A. (1978). Fuzzy Sets as Basis for a Theory of Possi-bility, *Fuzzy Sets and Systems*, Vol. 1, No. 1, pp. 3-28.