LUCIANO ALVES

EFICIÊNCIA DE MÉTODOS DE AGRUPAMENTO DE DADOS NA MODELAGEM NEBULOSA TAKAGI-SUGENO

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Engenharia de Produção e Sistemas.

CURITIBA

2007

LUCIANO ALVES

EFICIÊNCIA DE MÉTODOS DE AGRUPAMENTO DE DADOS NA MODELAGEM NEBULOSA TAKAGI-SUGENO

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Engenharia de Produção e Sistemas.

Área de Concentração: Automação e Controle de Processos.

Orientador: **Prof. Dr. Leandro dos Santos Coelho**

BANCA EXAMINADORA:

Prof. Dr. José Ernesto de Araújo Filho

Prof. Dr. Júlio César Nievola

Prof. Dr. Leandro dos Santos Coelho

CURITIBA

2007

Alves, Luciano
Eficiência de métodos de agrupamento de dados na modelagem nebulosa Takagi-Sugeno. Curitiba, 2007. 143p.

Dissertação – Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Engenharia de Produção e Sistemas.

1. Previsão de Séries Temporais 2. Sistemas Nebulosos 3. Lógica *Fuzzy* 4. Agrupamento de Dados 5. Agronegócios. Pontifícia Universidade Católica do Paraná. Centro de Ciências Exatas e de Tecnologia. Programa de Pós-Graduação em Engenharia de Produção e Sistemas.

Agradecimentos

Dedico aos meus pais, Marisa e Aparecido, que em suas simplicidades, me transmitiram o valor do caráter.

Dedico ao meu orientador, Leandro dos Santos Coelho, pelo apoio irrestrito, paciência e dedicação.

Dedico ao meu amigo, Carlos Alberto Campagnaro, pelo incentivo em todas as fases de estudo.

Em especial, aos meus filhos Nathalia e Matheus, cujo amor ilumina minha vida e a minha esposa Helena, companheira em todas as tristezas, alegrias e conquistas.

Agradeço aos membros da banca examinadora pela análise do trabalho e sugestões recebidas.

A todos que, direta ou indiretamente, colaboraram para realização deste trabalho.

A Deus, que tudo torna possível,...

“O analfabeto do século XXI não será aquele que não sabe ler e escrever, mas aquele que não consegue aprender, desaprender e aprender novamente”.

Alvin Tofler

“Todos os modelos são errados, mas alguns são úteis”.

George Box

RESUMO

A modelagem de sistemas dinâmicos não-lineares que usam sistemas nebulosos tem sido cada vez mais reconhecida como um paradigma da previsão de séries temporais. Modelar um sistema por meio de um sistema nebuloso engloba um conjunto de regras nebulosas e de funções de pertinências. Nesta dissertação, o problema de previsão de séries temporais é abordado. Nesse contexto, são dois os objetivos dessa dissertação. O primeiro é de abordar aspectos fundamentais de previsão de séries temporais. O outro é desenvolver modelos matemáticos a partir de dados observados, que permitam a construção de modelos nebulosos através de técnicas de agrupamento de dados. Além disso, deseja-se apresentar um estudo comparativo de desempenho dos algoritmos de agrupamento escolhidos, visando melhorar a eficiência de seus resultados em duas aplicações de previsão de séries temporais na área de agronegócios, abrangendo séries históricas do álcool hidratado e da soja.

Palavras-Chave: previsão de séries temporais, modelos nebulosos, agrupamento de dados, agronegócios.

ABSTRACT

The modeling of nonlinear dynamical systems that uses fuzzy systems has been increasingly recognized as paradigm of time series forecast. Modeling a system through a fuzzy system comprises a set of fuzzy rules and membership functions. In this dissertation, times series forecasting problem is approached. In this context, two objectives are presented. The first one is to approach the basic aspects of time series forecast. The other is to develop mathematical models from observed data's that allow building fuzzy models through data's clustering techniques. Also, want to present a comparative performance study of the chosen clustering algorithm, aiming at to improve the results efficiency of two time series forecasting applications in the agricultural business area, enclosing hydrated alcohol and soybean historical series.

Keywords: time series forecasting, fuzzy models, data's clustering, agricultural business.

Sumário

Índice de Figuras

Índice de Tabelas

Índice de Algoritmos

Lista de Abreviaturas

AD	*Partitioning Average Density*
AI	*Artificial Inteligence*
AES	Alisamento Exponencial Simples
ANP	Agência Nacional de Petróleo
AR	*AutoRegressive*
ARCH	*AutoRegressive Conditional Heterocedasticity*
ARMA	*AutoRegressive Moving Average*
ARIMA	*AutoRegressive Integrated Moving Average*
CBOT	*Chicago Board of Trade*
CEPEA	Centro de Estudos Avançados em Economia Aplicada
CML	*Classification Maximum Likelihood Algorithm*
DSP	*Digital Signal Processing*
ECG	Eletrocardiograma
EGARCH	*Exponential Generalized Autoregressive Conditional Heterocedasticity*
EM	*Expectation Maximization Algorithm*
FCA	*Fuzzy Clustering Algorithm*
FCM	*Fuzzy C-Means*
FCML	*Fuzzy Classification Maximum Likelihood*
FKM	*Fuzzy K-Means*
FOB	*Free on Board*
FS	*Fukuyama-Sugeno Index*
GA	*Genetic Algorithm*
GARCH	*Generalized Autoregressive Conditional Heterocedasticity*
GG	*Gath-Geva*
GIS	Sistemas de Informação Geográfica
GJR-GARCH	*Glosten-Jagannathan-Runkle GARCH*
GK	*Gustafson-Kessel*
GMM	*Gaussian Mixture Model*

HCM	*Hard C-Means*
HPV	Hipervolume Nebuloso
HW	Alisamento Exponencial Sazonal de Holt-Winters
LPV	*Linear Parameter Varying*
MA	*Moving Average*
MGG	*Modified Gath-Geva*
ML	*Maximum Likelihood Method*
MMS	Média Móvel Simples
MPL	*Maximum Penalized Likelihood*
MSE	*Mean Square Error*
PC	*Partitioning Coefficient*
PD	*Partitioning Density*
PE	*Partitioning Entropy*
PFCM	*Modified Penalty Fuzzy C-Means*
PVU	Posto-Veículo-Usina
PVD	Posto-Veículo-Destilaria
SVD	*Singular Value Decomposition*
TS	Takagi-Sugeno *Model*
USDA	*United States Department of Agriculture*
XB	*Xie-Beni Index*

Capítulo 1

INTRODUÇÃO

As técnicas da Inteligência Computacional – basicamente oriundas da teoria de Sistemas Nebulosos (*Fuzzy Systems*), Redes Neurais Artificiais, Inteligência Coletiva e Computação Evolutiva – são abordagens computacionais capazes de processar dados numéricos e/ou informações lingüísticas, cujos parâmetros podem ser ajustados a partir de exemplos (Coelho, 2003).

Recentemente é crescente o interesse no desenvolvimento de estratégias de identificação de sistemas não-lineares e mais especificamente em séries temporais. Tal fato é motivado por diversos fatores, tais como: (i) os avanços da teoria de sistemas não-lineares, que originaram metodologias de projeto aplicáveis a sistemas dinâmicos não-lineares complexos e (ii) o desenvolvimento de métodos eficazes em lidar com a existência de não-linearidades e incertezas em muitos dos sistemas dinâmicos presentes no meio industrial (Guerra, 2006) e, em particular, na área de agronegócios (Alves, 2002; Margarido *et al.* , 2002; Margarido e Sousa, 1998).

Uma série temporal consiste de um conjunto de observações no tempo e geralmente as observações em instantes de tempo próximos são correlacionadas. São objetivos do estudo de séries temporais, investigar o mecanismo gerador da série temporal, fazer previsões de valores futuros da série, descrever apenas o comportamento da série ou procurar periocidade relevante nos dados (Pucciarelli, 2005).

Neste contexto, os sistemas nebulosos têm sido utilizados em diferentes campos do conhecimento e em diversas aplicações com foco em previsão de séries temporais e identificação de sistemas.

1.1 MOTIVAÇÃO E RELEVÂNCIA DO TRABALHO

Com a necessidade da comunidade científica em lidar com sistemas cada vez mais complexos de serem modelados matematicamente, possibilitou a utilização de procedimentos de previsão de séries temporais. A implementação de sistemas inteligentes, principalmente a partir dos anos 70, vem de encontro a crescente necessidade de melhorar os procedimentos para a obtenção de modelos matemáticos representativos em termos de precisão e complexidade computacional (Passino e Yurkovich, 1998; Ribeiro e Meguelati, 2002).

A idéia de grupos de dados ou agrupamento possui uma representação similar à maneira humana de pensar – sempre que se lida com uma grande quantidade de dados, tende-se geralmente a sumarizar os dados em um pequeno número de grupos ou categorias, a fim de facilitar sua análise. O agrupamento divide um conjunto de dados em diversos grupos tais que a similaridade dentro de um grupo é maior do que aquela entre os grupos.

A abordagem comum de todas as técnicas de agrupamentos é de determinar os centros do grupo que representem cada grupo. O centro do grupo é uma referência para o algoritmo determinar se um ponto pertence ou não a um grupo.

Outro aspecto a ser considerado é a habilidade dos algoritmos de agrupamento de dados serem implementados de forma (i) *on-line* ou (ii) *off-line*. O agrupamento *on-line* é um processo em que cada vetor de entrada é usado para atualizar os centros do grupo de acordo com a posição do vetor. O sistema neste caso aprende onde os centros do grupo são introduzidos a cada nova entrada. No modo *off-line*, o sistema é apresentado com um treinamento de conjuntos de dados, na qual é usado para encontrar os centros do grupo, analisando todos os vetores de entrada no conjunto de treinamento. Uma vez que os centros do grupo são encontrados, após estes são fixados e usados *a priori* para classificar novos vetores de entrada.

Deve-se ressaltar que a determinação destes agrupamentos ou tentar categorizar os dados não são tarefas simples de serem executadas.

1.2 PROPOSTA DA DISSERTAÇÃO

A proposta dessa dissertação é de abordar aspectos relevantes em modelagem e previsão de séries temporais, levando-se em consideração dados do sistema dinâmico abordado e o projeto de modelos nebulosos baseados em métodos de agrupamento de dados. Desenvolver modelos matemáticos a partir de dados observados, com base nos casos estudados – análise de séries históricas dos preços do álcool hidratado e do grão de soja, e apresentar um estudo comparativo de desempenho de quatro métodos de agrupamento para o projeto de modelos nebulosos do tipo Takagi-Sugeno: (i) C-Médias Nebuloso, (ii) K-Médias Nebuloso, (iii) Gustafson-Kessel, (iv) Gath-Geva Modificado, (v) Modelo de Mistura Gaussiana (com Maximização da Esperança Esférica) e (vi) Modelo de Mistura Gaussiana (com Maximização da Esperança Elíptica). No caso específico desde trabalho, o principal objetivo de estudo de séries temporais é a previsão em curtíssimo prazo, ou seja, um passo à frente.

1.3 REVISÃO DA LITERATURA

Recentemente, o desenvolvimento de processos industriais modernos acarretou no surgimento de sistemas complexos tecnologicamente. Este desenvolvimento gerou a necessidade de pesquisas relativas a técnicas matemáticas, tais como abordagens da inteligência computacional – como os sistemas nebulosos, que possuam a capacidade de identificá-los e/ou concepção de métodos de previsão.

Com relação às pesquisas sobre sistemas nebulosos, essas se desenvolveram em duas principais direções. A primeira é uma abordagem lingüística ou qualitativa, na qual o sistema nebuloso é construído a partir de regras de produção (proposições). A segunda é uma abordagem quantitativa e está relacionada à teoria de sistemas clássicos e modernos. A combinação de informações qualitativas e quantitativas constitui a principal motivação para o uso de sistemas inteligentes, dando origem a várias contribuições sobre aspectos de flexibilidade, desempenho e robustez de algoritmos de previsão.

Em Vernieuwe *et al.* (2005) são analisados três métodos diferentes para construir modelos baseados em regras nebulosas do tipo Takagi-Sugeno usando particionamento em grades, método de agrupamento subtrativo e métodos de identificação com agrupamento de dados pelo método de Gustafson-Kessel. Os dados utilizados para parametrizar e validar os modelos consistem em registros de precipitação de hora em hora e descarte dos registros. Uma técnica modificada para predizer situações extremas e raras de níveis hidrométricos é apresentada em Luchetta e Manetti (2003). Uma aproximação usando sistemas nebulosos foi explorada para agrupamentos de dados e para o modelo de previsão de séries históricas. A metodologia foi desenvolvida, na colaboração com um fabricante italiano de equipamentos meteorológicos e ambientais, para o projeto de um sistema protótipo a ser instalado na bacia de *Padule di Fucecchio* no norte da Itália. A eficácia do método foi avaliada comparando-se o desempenho obtido com uma rede neural artificial.

Dois modelos que podem ser usados na previsão da demanda do turismo são apresentados por Wang (2004). Ambos os modelos são baseados em Inteligência Artificial (AI). Primeiramente, aplicou-se uma abordagem de rede neural artificial a previsão da demanda do turismo e testada empiricamente usando dados de Hong Kong do ano de 2000. Adicionalmente, evidências empíricas foram fornecidas com a utilização da teoria da caixa-cinza e de séries temporais usando modelos nebulosos.

Propõe-se em Wang *et al.* (2005) uma abordagem de aprendizagem de modelagem nebulosa baseada na aprendizagem competitiva nebulosa e decomposição em valores singulares (SVD, *Singular Value Decomposition*). Primeiramente, a aprendizagem competitiva nebulosa é usada para confirmar o espaço nebuloso das variáveis da entrada. Adicionalmente, os parâmetros conseqüentes do modelo nebuloso são determinados baseados no método SVD de mínimos quadrados recursivos. Para ilustrar o desempenho do método apresentado por Wang *et al.* (2005), simulações visando a predição da série temporal de Mackey-Glass são realizadas. Combinando cada entrada ou saída aprendida com o método proposto, o resultado simulado demonstra que a previsão das séries temporais caóticas de Mackey-Glass demonstra a eficácia desta abordagem.

Uma aproximação da mistura de agrupamento de dados é uma técnica importante na análise do conjunto de dados. Uma mistura de distribuições multinomial multivariável é usada geralmente para análise categórica dos dados com modelo latente da classe. Neste caso, a estimação de parâmetros é uma etapa importante para uma distribuição da mistura. São descritas por Yang e Yu (2005), quatro aproximações para estimar os parâmetros de uma distribuição multinomial multivariável: o método da probabilidade máxima (ML, *Maximum Likelihood*), algoritmo de maximização da expectativa (EM, *Expectation Maximization*), algoritmo da classificação de probabilidade máxima (CML, *Classification Maximum Likelihood*) e o método da classificação nebulosa de probabilidade máxima (FCML, *Fuzzy Classification Maximum Likelihood*). Lin *et al*. (2004) propuseram um algoritmo de agrupamento nebuloso (FCA) baseado na penalização da probabilidade máxima (MPL, *Maximum Penalized Likelihood*) e no modelo de penalidade do algoritmo C-Médias nebuloso (PFCM, *Modified Penalty Fuzzy C-Means*) para misturas normais. Lin *et al*. (2004) confirmaram com exemplos numéricos que o algoritmo de FCA-MPL (*Fuzzy Clustering Algorithm – Maximum Penalized Likelihood)* é mais eficiente computacionalmente do que o algoritmo EM.

Uma abordagem da metodologia de calibração automática para o modelo de Xianjiang é apresentada por Cheng *et al.* (2002), que consiste na calibração dos parâmetros de balanceamento da água e do parâmetro de roteamento de desligamento. Esse método é aplicado para resolver o problema de calibração, baseando-se na combinação dos modelos nebulosos com os algoritmos genéticos (GAs, *Genetic Algorithms*).

Uma arquitetura para operação de um forno de vidro reciclável é introduzida por Pina e Lima (2003). Esse modelo baseia-se em um esquema hierárquico visando à modelação e otimização do processo. A otimização do processo é realizada por um controlador baseado em um sistema especialista e usa algoritmos genéticos para resolver um problema de otimização multiobjetivo. A modelagem do processo é executada por um sistema de aprendizagem nebulosa.

Xu et *al.* (2004) apresenta uma abordagem de reconhecimento de padrões de combustível usando processamento de sinal digital (DSP, *Digital Signal Processing*) aliada a técnicas de inferência nebulosa. Um detector de flamas que contém três

fotodiodos é usado na derivação de sinais múltiplos que cobrem uma larga gama do espectro (do infravermelho ao ultravioleta) através da faixa visível. O processamento de sinal digital avançado e as técnicas de inferência nebulosa são desdobrados para identificar as impressões digitais no domínio do tempo e em freqüência, e finalmente identificar o tipo de carvão queimado sob circunstâncias constantes da combustão.

Döring *et al.* (2006) apresentam os fundamentos de diversos tipos de métodos de agrupamentos nebulosos, e visão geral de métodos de agrupamento nebulosos é fornecida, destacando-se as idéias subjacentes das diferentes aproximações. As partições dos conjuntos nebulosos são introduzidas com ênfase na interpretação de atribuições nebulosas e os graus de possibilidades do conjunto. São discutidas as classes de métodos baseados na função objetiva, a família de algoritmos de estimação alternada e a maximização da estimação nebulosa.

Devido a privatização da eletricidade em muitos países, a previsão de carga é crucial para o planejamento e operações na área de transmissão elétrica. Além disso, as previsões regionais de carga podem fornecer informações acuradas aos operadores da transmissão e distribuição de energia. É proposto por Pai (2006) o desenvolvimento de um sistema nebuloso elipsoidal híbrido para previsão da série temporal e aplicá-lo na previsão de cargas de eletricidade regionais em Taipei.

Um novo modelo para previsão de custo de inventário, baseado em modelos nebulosos e séries temporais, é proposto em Chen *et al.* (2007), na qual incorpora o conceito da seqüência Fibonacci, a estrutura do modelo proposto por Song e Chissom (1993, 1994) e o método de ponderações do modelo de Yu (2005).

Ainda pode-se citar, o uso de métodos nebulosos como ferramenta de modelos de identificação de sistemas não-lineares e multivariáveis quando da disponibilidade dos dados mensuráveis (Trabelsi *et al.*, 2004). Uma proposta de controlador adaptativo usando redes neurais e sistemas nebulosos foi apresentada em Gao e Joo (2003). Um método de aprendizagem híbrida para a construção de um classificador nebuloso baseado em cores de objetos presentes em imagens digitais coloridas é proposta em Bonventi (2005). A construção de uma matriz multi-setorial de insumo-produto em Minas Gerais para auxiliar a análise estratégica de políticas regionais de desenvolvimento e especificamente nas políticas de adensamento regional de cadeias

produtivas é elaborada por Simões (2003). A utilização de técnicas de aprendizado de máquina e análise de modelos que utilizam similaridade, árvores de decisão e modelos neuro-nebulosos na elaboração de modelos de análise de crédito para micro, pequena e média empresa são apresentadas por Souza *et al.* (2005). Uma aplicação de uma arquitetura em redes neurais para simulação presente e futura de comportamentos dinâmicos de uma série temporal de indicadores econômicos complexos no mercado norte-americano é proposta por Melin *et al.* (2007). Além disso, Landmann (2005) aborda o desenvolvimento e aplicação de uma metodologia heurística para programação e controle da produção na indústria de fundição.

Finalizando esta série de aplicações, Dote e Ovaska (2001) apresentam um apanhado de aplicações da inteligência computacional em diversas áreas, incluindo os modelos nebulosos, tais como: indústria aeroespacial, sistemas de comunicação, aplicações em consumo, sistemas de potência, engenharia de processos de manufatura e transportes.

1.4 ORGANIZAÇÃO DO TRABALHO

Este trabalho, fazendo uso das definições e considerações apresentadas neste capítulo, está organizado da seguinte forma.

O **Capítulo 2** descreve os conceitos básicos de previsão de séries temporais.

No **Capítulo 3** são abordados os conceitos de sistemas nebulosos e os procedimentos básicos a serem seguidos na previsão de séries temporais.

No **Capítulo 4** são discutidos os conceitos básicos sobre agrupamentos de dados e suas categorias, considerando os algoritmos: C-Médias Nebuloso, K-Médias Nebuloso, Gustafson-Kessel, Gath-Geva Modificado, Modelo de Mistura Gaussiana (com Maximização da Esperança Esférica) e Modelo de Mistura Gaussiana (com Maximização da Esperança Elíptica). Além disso, é abordado o método dos mínimos quadrados e a estimação dos parâmetros do conseqüente do modelo nebuloso do tipo Takagi-Sugeno (TS).

No **Capítulo 5** são descritas as séries históricas dos preços do álcool hidratado e do grão de soja.

O **Capítulo 6** apresenta os resultados das simulações de cada sistema e uma análise de resultados obtidos. Neste contexto, um estudo comparativo de técnicas de agrupamento de dados é apresentado.

O **Capítulo 7** enfoca as conclusões e propostas para trabalhos futuros.

Capítulo 2

SÉRIES TEMPORAIS

Uma série temporal é uma coleção de observações realizadas seqüencialmente e ordenadas ao longo do tempo, dada por $z_t, t = 1,2,\ldots,n$, podendo ser discretas (como os valores mensais de temperatura) ou contínuas (como o registro de um eletrocardiograma - ECG). Esse conjunto é uma parte de uma trajetória dentre muitas que poderiam ter sido observadas (Morettin e Toloi, 1987).

2.1 CONSIDERAÇÕES INICIAIS

Segundo Coelho (2000), os principais ingredientes ao problema da previsão de séries temporais são: um conjunto de dados de entradas e saídas; uma classe de modelos e suas estruturas; um critério de adequação entre os dados e o modelo; as rotinas para validação e aceitação dos modelos resultantes. Com a definição dessas questões, deve-se proceder à resolução das seguintes etapas: a detecção de não-linearidades, a escolha da estrutura, a estimação dos parâmetros, a validação do modelo e a previsão (a curto ou em longo prazo). Como a maior parte dos procedimentos estatísticos foi desenvolvida para analisar observações independentes, o estudo de séries temporais requer o uso de técnicas específicas (Morettin e Toloi,1987).

Durante as décadas de 30 e 40, considerava-se como representação das séries temporais a combinação de quatro componentes não observáveis e distintas (Abelém, 1994; Purcote, 2006): t_t (Tendência), c_t (Cíclica), s_t (Sazonalidade), e_t (Erro ou Ruído Aleatório); ou seja, $z_t = f(t_t, s_t, c_t, e_t)$.

Os componentes de tendência são, freqüentemente, aquelas que produzem mudanças graduais e regulares ao longo do tempo, como por exemplo, índice de mortalidade infantil, índice de alfabetização, entre outros.

As componentes cíclicas são aquelas que provocam oscilações de subida e de queda nas séries de forma suave, repetitiva e forma irregular ao longo da componente de tendência, como por exemplo, índices de demanda do produto, comportamento do mercado financeiro, entre outros.

Assim como as componentes cíclicas, as componentes sazonais também provocam oscilações de subida e de queda nas séries, porém de forma regular e com movimentos previsíveis ao longo da componente de tendência, como por exemplo, ano a ano, mês a mês, semana a semana, dia a dia, entre outros.

O componente erro ou ruído aleatório representa movimentos ascendentes e descendentes da série e aparecem como flutuações de período curto e com deslocamento inexplicável, após a ocorrência de um efeito de tendência, cíclico ou sazonal.

As componentes são decompostas em três modelos: (i) aditivo $(z_t = t_t + c_t + s_t + e_t)$, (ii) multiplicativo $(z_t = t_t.c_t.s_t.e_t)$ e (iii) misto $(z_t = t_t.c_t.s_t + e_t)$.

No caso específico desde trabalho, o principal objetivo de estudo de séries temporais é a previsão em curtíssimo prazo (um passo à frente).

2.2 MODELOS E PROCEDIMENTOS DA PREVISÃO DE SÉRIES TEMPORAIS

Um modelo que descreve uma série necessariamente não conduz a um procedimento de previsão, sendo necessário especificar uma função perda mais apropriado, como o erro quadrático médio de previsão dado por

$$e\left[z(t+h) - \hat{z}_t(h)\right]^2 \tag{2.1}$$

onde $\hat{z}_t(h)$ é a previsão de $z(t+h)$, de tempo t e horizonte de previsão h.

Os modelos para séries temporais podem ser classificados em paramétricos, tais como os modelos de regressão de erros, os modelos auto-regressivos de médias móveis (ARMA, *AutoRegressive Moving Average*) e os modelos auto-regressivos integrados de médias móveis (ARIMA, *AutoRegressive Integrated Moving Average*) e os não paramétricos, tais como a função de autocovariância ou função de autocorrelação e a sua respectiva transformada de Fourier.

2.3 MEDIDAS DE ACUIDADE DO MÉTODO DE PREVISÃO DE SÉRIES TEMPORAIS

De acordo com Mueller (1996), a suposição básica de qualquer técnica de previsão de séries temporais é que o valor observado na série fica determinado por um padrão que se repete no tempo e por alguma influência aleatória. Isto significa dizer que mesmo quando o padrão exato que caracteriza o comportamento da série temporal tenha sido isolado, algum desvio ainda existirá entre os valores da previsão e os valores realmente observados. Essa aleatoriedade não pode ser prevista; entretanto, se isolada, sua magnitude pode ser estimada e usada para determinar a variação ou erro entre as observações e previsões realizadas.

Segundo Mueller (1996), a acuidade de um método de previsão pode ser mensurada através de muitas medidas de erro. Dessa forma, a verificação da adequação de um determinado modelo matemático, supostamente representativo da série histórica de dados, é dependente da medida de erro adotada para efetuar essa validação.

2.4 MODELOS LINEARES

2.4.1 Métodos Simples de Previsão de Séries Temporais

Os métodos assim classificados efetuam a previsão do valor futuro da série temporal pelo alisamento das observações passadas da série de interesse. Assumindo que os

valores extremos da série representam flutuações aleatórias, o propósito desses métodos consiste em identificar o padrão básico presente nos dados históricos e, então, usar esse padrão para prever valores futuros. Entre os métodos simples de previsão destacam-se os da Média Móvel Simples (MMS), o Alisamento Exponencial Simples (AES), o Alisamento Exponencial de Holt e o Alisamento Exponencial Sazonal de Holt-Winters (HW).

(i) Modelos para Séries Estacionárias: Segundo Morettin e Toloi (1987), um processo estacionário se desenvolve no tempo aleatoriamente ao redor de uma média constante, de modo que a escolha de uma origem no tempo (tempo inicial) não seja relevante, refletindo alguma forma de equilíbrio estável. Assumindo que o modelo é adequado para descrever as observações $z_1,\ldots,z_n$, as séries temporais podem ser decompostas em nível e ruído aleatório, dadas por

$$z_t = \mu_t + e_t,\ t = 1,\ldots,n \tag{2.2}$$

onde z_t é um processo estocástico estacionário, μ_t o nível da série e e_t um ruído aleatório. Partindo-se desse pressuposto, são descritos dois modelos de previsão que possam representar a parte determinística e seu padrão de comportamento.

- Média Móvel Simples (MMS): Os modelos de médias móveis são adequados quando se trabalha com hipótese de estacionaridade na série, sem que se identifique uma tendência de aumento ou decréscimo nos dados observados. Este modelo consiste em calcular a média aritmética das φ observações mais recentes registradas, ou seja,

$$m_t = \frac{z_t + z_{t-1} + \cdots + z_{t-\varphi+1}}{\varphi} \tag{2.3}$$

$$m_t = m_{t-1} + \left(\frac{z_t + z_{t-\varphi}}{\varphi} \right) \tag{2.4}$$

O termo "média móvel" é atribuído porque a cada período, a observação mais antiga é substituída pela mais recente e uma nova média é calculada. A previsão de todos os valores futuros da série a partir da origem t é dada pela última média calculada, isto é,

$$\hat{z}_t(h) = m_t \tag{2.5}$$

Para todo horizonte de previsão $h = 1,2,3,\ldots,$ pode-se escrever:

$$\hat{z}_t(h) = \hat{z}_{t-1}(h+1) + \left(\frac{z_t + z_{t-\varphi}}{\varphi} \right) \tag{2.6}$$

A equação (2.6) pode ser interpretada como um mecanismo de atualização de previsão, pois a cada instante ou a cada observação, corrige a estimativa prévia de z_{t+h}.

A aplicação do método de MMS depende do número φ de observações utilizadas na média. Um valor pequeno de φ implica numa reação mais rápida. Dois casos extremos são: (i) $\varphi = 1$, então o valor mais recente da série é utilizado como previsão de todos os valores futuros. Este é o tipo de previsão mais simples, denominado "método ingênuo"; (ii) $\varphi = n$, então a previsão será igual à média aritmética de todos os dados observados. Esse caso é indicado quando a série é altamente aleatória (se a aleatoriedade de e_t predominar sobre a mudança de nível).

Assim o valor de φ deve ser proporcional à aleatoriedade de e_t. Um procedimento objetivo é selecionar o valor de φ que fornece a "melhor previsão" a um passo das observações já obtidas, ou seja, encontrar o valor de φ que minimize

$$s = \sum_{t=l+1}^{n} \left(z_t - \hat{z}_{t-1}(\varphi) \right)^2 \tag{2.7}$$

onde l é escolhido de tal modo que o valor inicial utilizado na equação (2.4) não influencie a previsão.

• Alisamento Exponencial Simples (AES): A princípio, o método conhecido como Alisamento Exponencial Simples se assemelha ao da Média Móvel por extrair das observações da série temporal o comportamento aleatório pelo alisamento dos dados históricos. Entretanto, a inovação introduzida pelo Alisamento Exponencial Simples advém do fato de este método atribuir ponderações diferentes às observações mais recentes da série. Enquanto que na Média Móvel as observações usadas para encontrar a previsão do valor futuro contribuem em igual proporção para o cálculo dessa previsão. No Alisamento Exponencial Simples é uma média ponderada que atribui ponderações maiores às observações mais recentes. Este método utiliza todos os valores históricos, com coeficientes de ponderação que decrescem exponencialmente. Outro argumento para o tratamento diferenciado das observações da série temporal é fundamentado na suposição de que as últimas observações contêm mais informações sobre o futuro e, portanto, são mais relevantes à previsão. O modelo é descrito pela equação:

$$\bar{z}_t = \alpha z_t + (1-\alpha)\bar{z}_{t-1} \tag{2.8}$$

$$\bar{z}_0 = z_1, \ t = 1,\ldots,n \tag{2.9}$$

onde $\bar{z}_t$ é denominado valor exponencial alisado e α é a constante de alisamento. O último valor exponencialmente alisado é utilizado na previsão de todos os valores futuros da série, dados por

$$\hat{z}_t(h) = \bar{z}_t, \ h = 1,2,3,\ldots, \tag{2.10}$$

$$\hat{z}_t = \alpha z_t + (1-\alpha)\hat{z}_{t-1}(h+1) \tag{2.11}$$

Dessa forma, o problema de armazenamento de observações é reduzido pela aplicação da equação (2.11), pois a previsão pode ser analisada utilizando apenas a

observação mais recente z_t, a previsão imediatamente anterior $\hat{z}_{t-1}$ e o valor de α. Para eliminar uma das desvantagens do método MMS em relação a atribuição de ponderações às φ observações, utiliza-se recursivamente a equação (2.8), dessa forma tem-se

$$\bar{z}_t = z_t + \alpha(1-\alpha)\bar{z}_{t-1} + +\alpha(1-\alpha)^2\bar{z}_{t-2} + \cdots + +\alpha(1-\alpha)^n\bar{z}_{t-n} \qquad (2.12)$$

Quanto menor o valor da constante de alisamento α, mais estáveis são as previsões, visto que a utilização de valores baixos para α implica na atribuição de ponderações maiores às observações passadas e, conseqüentemente, qualquer flutuação aleatória no presente contribui com menor importância para a obtenção da previsão. Contudo, não há uma metodologia que oriente quanto à seleção de um valor apropriado para α, sendo normalmente encontrado por tentativa e erro. Um procedimento mais objetivo seria a seleção do valor de α que forneça a "melhor previsão das observações contidas na série temporal".

Segundo Morettin e Toloi (1987), o método de Alisamento Exponencial Simples é muito utilizado devido ser de fácil entendimento e flexibilidade permitida pela variação de α, além da necessidade de armazenar somente z_t, $\bar{z}_t$ e α. Sua principal desvantagem é a dificuldade em determinar o valor mais apropriado da constante de alisamento α.

(ii) Modelos de Alisamento Exponencial para Séries que apresentam Tendência: Assumindo que o modelo adequado para descrever as observações $z_1,\ldots,z_n$, as séries temporais não sazonais, compostas localmente da soma de nível da série μ_t, tendência t_t e ruído aleatório e_t, com média zero e variância constante, é dada por

$$z_t = \mu_t + t_t + e_t,\ t = 1,\ldots,n \qquad (2.13)$$

• Método de Holt: Este método modifica a equação (2.12) pela incorporação na tendência t_t de longo prazo, sendo a base do método de alisamento exponencial com dois parâmetros de Holt (α, β). O método de Holt serve como uma abordagem de previsão para séries que exibem tendência linear, calculando uma estimativa do nível de série μ_t e a tendência t_t (Rocha, 2003). A função previsão para um número de passos à frente h ficaria, então, da seguinte forma

$$\hat{z}_t(h) = \bar{z}_t - h\hat{t}_t, \ h = 1,2,3,\ldots, \tag{2.14}$$

$$\bar{z}_t = \alpha z_t + (1-\alpha)\bar{z}_{t-1}\left(\bar{z}_{t-1} - \hat{t}_{t-1}\right), \ 0 < \alpha < 1 \text{ e } t = 2,\ldots,n \tag{2.15}$$

$$\hat{t}_t = \beta\left(\bar{z}_t - \bar{z}_{t-1}\right) + (1-\beta)\hat{t}_{t-1}, \ 0 < \beta < 1 \text{ e } t = 2,\ldots,n \tag{2.16}$$

As equações (2.13) e (2.16) podem ser utilizadas na atualização da previsão para z_{t+1}, dessa forma

$$\bar{z}_{t+1} = \alpha z_{t+1} + (1-\alpha)\left(\bar{z}_t - \hat{t}_t\right) \tag{2.17}$$

$$\hat{t}_{t+1} = \beta\left(\bar{z}_{t+1} - \bar{z}_t\right) + (1-\beta)\hat{t}_t \tag{2.18}$$

e a nova previsão para z_{t+h} é dada por

$$\hat{z}_{t+1}(h-1) = \bar{z}_{t+1} + (h-1)\hat{t}_{t+1} \tag{2.19}$$

Os valores de alisamento α e β são determinados de forma que sejam minimizadas as somas dos erros quadráticos de previsão um-passo-à-frente, com o erro em cada período t sendo determinado por

$$r_t = z_t - \hat{z}_t = z_t - \left(\bar{z}_{t-1} + \hat{t}_{t-1}\right) \tag{2.20}$$

A soma dos erros quadráticos dos distintos períodos t é determinada por

$$\sum_{t=1}^{n} r_t^2 = \sum_{t=1}^{n} \left(z_t - \hat{z}_t \right)^2 \tag{2.21}$$

(iii) Modelos de Alisamento Exponencial para Séries Sazonais: Existem dois tipos de procedimentos cuja utilização depende das características da série considerada. Tais procedimentos são baseados em três equações com constantes de alisamento diferentes, que são associadas às componentes de nível, tendência e sazonalidade da série temporal.

- Alisamento Exponencial Sazonal de Holt-Winters (HW): Este método é adequado para séries que apresentam, além da tendência, uma componente sazonal. Existem dois tipos de procedimentos cuja utilização depende das características da série considerada.

A incorporação do efeito sazonal pode ser feita da forma aditiva ou multiplicativa. Como cada uma destas formas pode fornecer previsões distintas, a opção dever ser feita em função das características da série temporal z_t (Rocha, 2003). A variedade mais usual do método HW considera o fator sazonal f_t, como sendo multiplicativo, enquanto a tendência permanece aditiva, isto é

$$z_t = \mu_t f_t + t_t + e_t, \; t = 1, \ldots, n \tag{2.22}$$

O procedimento anterior pode ser modificado para tratar de situações onde o fator sazonal f_t é aditivo, isto é

$$z_t = \mu_t + f_t + t_t + e_t, \; t = 1, \ldots, n \tag{2.23}$$

As vantagens são semelhantes às da utilização do método de Holt, sendo que o método de HW é adequado à análise de séries com padrão de comportamento mais geral. As desvantagens são as dificuldades em determinar os valores apropriados das

constantes de suavização (α, β, γ). Para isso é realizada de modo a tornar mínima a soma dos quadrados dos erros de ajustamento.

1) Série Sazonal Multiplicativa: Para demonstração do método sazonal multiplicativo de HW, considere que s representa o número de períodos sazonais. A função previsão apresenta a seguinte forma

$$\hat{z}_t(h) = \left(\bar{z}_t - h\hat{t}_t\right) f_{t+h-s}\,,\ h = 1,2,\ldots,s \tag{2.24}$$

$$\hat{z}_t(h) = \left(\bar{z}_t - h\hat{t}_t\right) f_{t+h-2s}\,,\ h = s+1,\ldots,2s \tag{2.25}$$

em que

$$\hat{f}_t = \gamma\left(\frac{z_t}{\bar{z}_t}\right) + (1-\gamma)\hat{f}_{t-s}\,,\ 0 < \gamma < 1 \text{ e } t = s+1,\ldots,n \tag{2.26}$$

$$\bar{z}_t = \alpha\left(\frac{z_t}{\hat{f}_{t-s}}\right) + (1-\alpha)\left(\bar{z}_{t-1} + \hat{t}_{t-1}\right),\ 0 < \alpha < 1 \text{ e } t = s+1,\ldots,n \tag{2.27}$$

$$\hat{t}_t = \beta\left(\bar{z}_t - \bar{z}_{t-1}\right) + (1-\beta)\hat{t}_{t-1}\,,\ 0 < \beta < 1 \text{ e } t = s+1,\ldots,n \tag{2.28}$$

As equações (2.26), (2.27) e (2.28) podem ser utilizadas na atualização da previsão para z_{t+1}, dessa forma

$$\hat{f}_{t+1} = \gamma\left(\frac{z_{t+1}}{\bar{z}_{t+1}}\right) + (1-\gamma)\hat{f}_{t+1-s} \tag{2.29}$$

$$\bar{z}_{t+1} = \alpha\left(\frac{z_{t+1}}{\hat{f}_{t+1-s}}\right) + (1-\alpha)\left(\bar{z}_t - \hat{t}_t\right) \tag{2.30}$$

$$\hat{t}_{t+1} = \beta\left(\bar{z}_{t+1} - \bar{z}_t\right) + (1-\beta)\hat{t}_t \tag{2.31}$$

e a nova previsão para z_{t+h} é dada por

$$\hat{z}_{t+1}(h-1) = \left(\bar{z}_{t+1} + (h-1)\hat{t}_{t+1}\right)\hat{f}_{t+1+h-s}\,,\ h = 1,2,\ldots,s+1 \tag{2.32}$$

$$\hat{z}_{t+1}(h-1) = \left(\bar{z}_{t+1} + (h-1)\hat{t}_{t+1}\right)\hat{f}_{t+1+h-2s}\,,\ h = s+2,\ldots,2s+1 \tag{2.33}$$

Os valores iniciais das equações recorrentes são calculados por meio das seguintes fórmulas:

$$\hat{f}_j = \frac{z_j}{\left(\frac{1}{s}\right)\sum_{k=1}^{s} z_k}, \ j = 1,2,\ldots,s \tag{2.34}$$

$$\bar{z}_s = \frac{1}{s}\sum_{k=1}^{s} z_k \tag{2.35}$$

$$\hat{t}_s = 0 \tag{2.36}$$

2) Série Sazonal Aditiva: Para demonstração do método sazonal aditivo de HW, considere que s representa o número de períodos sazonais. A função previsão apresenta a seguinte forma

$$\hat{z}_t(h) = \bar{z}_t + h\hat{t}_t + f_{t+h-s}, \ h = 1,2,\ldots,s \tag{2.37}$$

$$\hat{z}_t(h) = \bar{z}_t + h\hat{t}_t + f_{t+h-2s}, \ h = s+1,\ldots,2s \tag{2.38}$$

em que

$$\hat{f}_t = \gamma(z_t - \bar{z}_t) + (1-\gamma)\hat{f}_{t-s}, \ 0 < \gamma < 1 \text{ e } t = s+1,\ldots,n \tag{2.39}$$

$$\bar{z}_t = \alpha\left(z_t - \hat{f}_{t-s}\right) + (1-\alpha)\left(\bar{z}_{t-1} + \hat{t}_{t-1}\right), \ 0 < \alpha < 1 \text{ e } t = s+1,\ldots,n \tag{2.40}$$

$$\hat{t}_t = \beta(\bar{z}_t - \bar{z}_{t-1}) + (1-\beta)\hat{t}_{t-1}, \ 0 < \beta < 1 \text{ e } t = s+1,\ldots,n \tag{2.41}$$

As equações (2.39), (2.40) e (2.41) podem ser utilizadas na atualização da previsão para z_{t+1}, dessa forma

$$\hat{f}_{t+1} = \gamma(z_{t+1} - \bar{z}_{t+1}) + (1-\gamma)\hat{f}_{t+1-s} \tag{2.42}$$

$$\bar{z}_{t+1} = \alpha\left(z_{t+1} - \hat{f}_{t+1-s}\right) + (1-\alpha)\left(\bar{z}_t - \hat{t}_t\right) \tag{2.43}$$

$$\hat{t}_{t+1} = \beta(\bar{z}_{t+1} - \bar{z}_t) + (1-\beta)\hat{t}_t \tag{2.44}$$

e a nova previsão para z_{t+h} é dada por

$$\hat{z}_{t+1}(h-1) = \bar{z}_{t+1} + (h-1)\hat{t}_{t+1} + \hat{f}_{t+1+h-s},\ h = 1,2,\ldots,s+1 \tag{2.45}$$

$$\hat{z}_{t+1}(h-1) = \bar{z}_{t+1} + (h-1)\hat{t}_{t+1} + \hat{f}_{t+1+h-2s},\ h = s+2,\ldots,2s+1 \tag{2.46}$$

Tanto para o método aditivo como para o multiplicativo, os valores de alisamento α, β e γ são determinados de forma que sejam minimizadas as somas dos erros quadráticos de previsão um passo à frente, com o erro em cada período t sendo determinado por

$$r_t = z_t - \hat{z}_t = z_t - \left(\bar{z}_{t-1} + \hat{t}_{t-1} + \hat{f}_{t-s}\right) \tag{2.47}$$

A soma dos erros quadráticos dos distintos períodos t é determinada pela equação (2.21).

2.4.2 Outros Métodos de Previsão de Séries Temporais

No universo dos métodos de previsão de séries temporais mais aprimorados encontram-se o modelo Autoregressivo (AR, *AutoRegressive*), o modelo de Médias Móveis (MA, *Moving Average*), o modelo Autoregressivo de Médias Móveis (ARMA), o modelo Autoregressivo Integrado de Médias Móveis (ARIMA), o modelo Autoregressivo Heterocedástico Condicional (ARCH, *AutoRegressive Conditional Heterocedasticity*), o modelo Autoregressivo Heterocedástico Condicional Generalizado (GARCH, *Generalized AutoRegressive Conditional Heterocedasticity*), o modelo Autoregressivo Heterocedástico Condicional Generalizado Exponencial (EGARCH, *Exponential Generalized AutoRegressive Conditional Heterocedasticity*), o modelo Autoregressivo Heterocedástico Condicional Generalizado de Glosten, Jagannathan e Runkle (GJR–GARCH), entre outros (Bollerslev, 1986; Wang, 2004; Gooijer e Hyndman, 2006). Os

métodos assim classificados obtêm a previsão de algum valor futuro da série temporal, pela combinação dos valores reais passados e/ou dos erros ocorridos.

(i) Modelo Autoregressivo (AR): O método de Box e Jenkins (Box e Jenkins, 1970) permite prognosticar séries temporais x_t, descritas somente por seus valores e termos aleatórios, é representado por

$$x_t = \phi_1 x_{t-1} + e_t \tag{2.48}$$

onde x_t corresponde à observação da série temporal no tempo t, ϕ corresponde ao parâmetro de ajuste do modelo AR e e_t é um ruído aleatório.

A abordagem do método de Box e Jenkins é de efetuar um ajuste com modelagem estocástica (probabilística) em séries estacionárias (sem tendência definida), constituindo-se em três estágios principais (Martin, 2000): (i) Identificação – a estacionaridade é investigada por meio de uma função de autocorrelação e é efetuada uma tentativa de avaliação a construção de ordem de um modelo; (ii) Estimação dos Parâmetros – a melhor maneira de estimar é através da utilização de procedimentos de otimização não linear; e (iii) Diagnóstico – os resíduos, a periodicidade e os erros dos estimadores são examinados, para evitar valores sobreestimados.

Estando os modelos identificados, estimados e diagnosticados, cumpre-se a etapa da previsão de valores futuros. Uma importante utilização destas classes de modelos seja, a previsão dos valores da seqüência x_t (Rocha, 2003).

Para que o modelo auto-regressivo x_t seja francamente estacionário a sua variância deve exibir a propriedade de constância ao longo do tempo t. Além disso, deve apresentar covariâncias independentes do tempo t e dependentes somente de k, que é a distância que separa as observações. Dessa forma, a variância e a covariância de AR são, respectivamente,

$$\gamma_0 = \phi^2 \gamma_0 + \sigma_e^2 \tag{2.49}$$

e

$$\gamma_k = \phi^k \gamma_0 \tag{2.50}$$

É possível prognosticar o valor futuro da série em análise, pela representação das observações da série temporal pela equação (2.48), determinando a ordem do modelo e os parâmetros estimados. Generalizando a equação (2.48), o modelo auto-regressivo AR(p) será o resultado da soma ponderada de seus p valores passados e do ruído e_t, representado por

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + e_t \tag{2.51}$$

onde x_t corresponde à observação da série temporal no tempo t, ϕ_p corresponde ao parâmetro do modelo AR de ordem p e e_t é um ruído branco normalmente distribuído com média zero e variância constante σ^2.

Dessa forma, a variância e a covariância de AR(p) são, respectivamente,

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \cdots + \phi_p \gamma_p + \sigma_e^2 \tag{2.52}$$

e

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p} \tag{2.53}$$

(ii) Modelo de Médias Móveis (MA): A série temporal é resultante da combinação linear dos termos aleatórios em t e em períodos anteriores, sendo representado por

$$x_t = e_t - \theta e_{t-1} \tag{2.54}$$

onde x_t corresponde à observação da série temporal no tempo t, θ corresponde ao parâmetro de ajuste do modelo MA e e_t é um ruído aleatório. Dessa forma, a variância e a covariância de MA são dadas pelas seguintes equações,

$$\gamma_0 = \sigma_e^2 + \theta^2\sigma_e^2 = (1+\theta^2)\sigma_e^2 \tag{2.55}$$

e

$$\gamma_k = -\theta_k\sigma_e^2 \tag{2.56}$$

É possível prognosticar o valor futuro da série em análise, pela representação das observações da série temporal pela equação (2.54) que é similar à equação (2.48), exceto pelo fato de que o valor previsto às observações depende dos valores dos erros observados em cada período passado, ao invés das observações propriamente ditas. Generalizando a equação (2.54), o modelo de Médias Móveis MA(q) será o resultado da soma ponderada de seus q valores passados e do ruído e_t, representado por

$$x_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \tag{2.57}$$

onde x_t corresponde à observação da série temporal no tempo t, θ_q corresponde ao parâmetro do modelo MA de ordem q e e_t é um ruído branco normalmente distribuído com média zero e variância constante σ^2.

Dessa forma, a variância e a covariância de MA(q) são regidas, respectivamente, pelas seguintes equações,

$$\gamma_0 = (1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2)\sigma_e^2 \tag{2.58}$$

e

$$\gamma_k = (-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots\theta_{q-k}\theta_q)\sigma_e^2 \tag{2.59}$$

(iii) Modelo Auto-regressivo e de Médias Móveis (ARMA): A série temporal é função de seus valores históricos e pelos termos aleatórios em t e em período corrente e passados. Em sua versão mais simples é uma combinação dos modelos AR e MA, representado por

$$x_t = \phi\, x_{t-1} + e_t - \theta\, e_{t-1} \tag{2.60}$$

onde x_t corresponde à observação da série temporal no tempo t, (ϕ,θ) correspondem aos parâmetros de ajustes do modelo ARMA e e_t é um ruído aleatório. Dessa forma, a variância e a covariância de ARMA são respectivamente:

$$\gamma_0 = \frac{\left(\sigma_e^2 + \theta^2\sigma_e^2 - 2\phi\theta\sigma_e^2\right)}{1-\phi^2} = \frac{\left(1+\theta^2-2\phi\theta\right)\sigma_e^2}{1-\phi^2} \tag{2.61}$$

e

$$\gamma_k = \phi\gamma_{k-1} \tag{2.62}$$

Generalizando a equação (2.60), é possível verificar que o modelo auto-regressivo de médias Móveis ARMA(p,q) relaciona os valores futuros com as observações passadas, assim como também com os erros passados apurados entre os valores reais e previstos, representado por

$$x_t = \phi_1\, x_{t-1} + \phi_2\, x_{t-2} + \cdots + \phi_p\, x_{t-p} + e_t - \theta_1\, e_{t-1} - \theta_2\, e_{t-2} - \cdots - \theta_q\, e_{t-q} \tag{2.63}$$

onde x_t corresponde à observação da série temporal no tempo t, (ϕ_p,θ_q) correspondem aos parâmetros do modelo ARMA de ordem (p,q) e e_t é um ruído branco normalmente distribuído com média zero e variância constante σ^2.

Mediante o operador de defasagem b, a equação (2.63) pode ser escrita como,

$$x_t = \left(\phi_1 b + \phi_2 b^2 + \cdots + \phi_p b^p\right) x_t + \left(1 - \theta_1 b - \theta_2 b^2 - \cdots - \theta_q b^q\right) e_t \quad (2.64)$$

$$\left(1 - \phi_1 b - \phi_2 b^2 - \cdots - \phi_p b^p\right) x_t = \left(1 - \theta_1 b - \theta_2 b^2 - \cdots - \theta_q b^q\right) e_t \quad (2.65)$$

$$\text{ou simplesmente, } \phi(b) x_t = \theta(b) e_t \quad (2.66)$$

(iv) Modelo Auto-regressivo Integrado e de Médias Móveis (ARIMA): Quando a série temporal analisada não satisfazer a condição de estacionaridade, a série temporal x_t é diferenciada d vezes até que esta condição se torne válida, obtendo-se uma série estacionária. Com a série derivada deste processo, descreve-se como um modelo ARMA(p,q). Em seguida, a ordem do processo ARMA é identificada pela análise dos coeficientes de autocorrelação. Ainda nesta etapa são efetuadas estimativas preliminares dos parâmetros do modelo identificado. Neste contexto, o objetivo da identificação é determinar os valores de (p,d,q) do modelo ARIMA(p,d,q). Genericamente, o modelo é representado por

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \cdots + \phi_p w_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (2.67)$$

$$w_t = x_t - x_{t-d} = \Delta^d x_t \quad (2.68)$$

onde x_t corresponde à observação da série temporal no tempo t, (ϕ_p, θ_q) correspondem aos parâmetros do modelo ARMA de ordem (p,q), e_t é um ruído branco normalmente distribuído com média zero e variância constante σ^2 e d equivale ao grau de homogeneidade não estacionária.

Após a identificação do modelo que seja uma representação adequada do mecanismo gerador da série, a estimação dos parâmetros desse modelo é efetuada. Estimado o modelo, a verificação de sua habilidade em representar os fenômenos observáveis da série temporal é confirmada pela análise dos erros do modelo proposto. Caso a inadequação fique evidenciada, o ciclo de identificação, estimação e verificação são novamente aplicados, até que a representação apropriada seja encontrada. Após a validação do modelo, a previsão dos valores futuros da série temporal modelada pode

ser obtida. Mediante o operador de defasagem b, a equação (2.67) pode ser escrita como

$$\left(1-\phi_1 b-\phi_2 b^2-\cdots-\phi_p b^p\right)w_t=\left(1-\theta_1 b-\theta_2 b^2-\cdots-\theta_q b^q\right)e_t \tag{2.69}$$

$$\text{ou simplesmente } \phi(b)w_t=\theta(b)e_t \tag{2.70}$$

$$\text{sendo } w_t=(1-b)^d x_t \tag{2.71}$$

$$\text{tem-se } (1-b)^d\phi(b)x_t=\theta(b)e_t \tag{2.72}$$

(v) Modelo Autoregressivo Heterocedástico Condicional (ARCH): os modelos mais simples não consideravam o fato de a volatilidade variar com o tempo. A classe de modelos ARCH, descreve a variância heterocedástica condicional, ou seja, não é constante ao longo do tempo. Neste modelo, a variância é uma função determinística (tipicamente quadrática) das inovações passadas (Bollerslev, 1986; Gooijer e Hyndman, 2006). A representação do modelo ARCH$(1,1)$ é dada por

$$\sigma_t^2=\alpha_0+\alpha_1 x_{t-1}^2 \tag{2.73}$$

onde α_0 e α_1 são constantes reais e positivas.

Para o modelo ARCH(p) ser bem definido e a variância condicional ser positiva, as restrições paramétricas devem satisfazer $\alpha_0>0$ e $\alpha_i>0$, $i=1,2,\ldots,p$, tal que

$$\sigma_t^2=\alpha_0+\sum_{i=1}^{p}\alpha_i x_{t-i}^2 \tag{2.74}$$

onde os parâmetros $\alpha_0,\alpha_1,\ldots,\alpha_p$ são constantes positivas e x_t é uma variável randômica, com média zero e variância σ_t^2, gerada a partir de uma função densidade de probabilidade $P_t(x_t)$ prévia e arbitrariamente escolhida. Uma vez escolhidos os

parâmetros $\alpha_0, \alpha_1, \ldots, \alpha_p$ e a forma de $P_t(x_t)$, o processo ARCH(p) está definido. Itera-se a equação (2.74) e a distribuição $P(x)$ é determinada (Favaro, 2007).

(i) Modelo Autoregressivo Heterocedástico Condicional Generalizado (GARCH): uma importante extensão do modelo ARCH é a sua versão generalizada proposta por Bollerslev (1986), denominada GARCH. Neste modelo, a função linear da variância condicional inclui também variâncias passadas. Assim sendo, a volatilidade dos retornos depende dos quadrados dos erros anteriores e também de sua própria variância em momentos anteriores. O modelo GARCH$(1,1)$ fornece uma boa parametrização para retornos diários e é mais utilizada em séries financeiras. Entretanto, a qualidade da previsão fora da amostragem sugere que os resultados devem ser examinados com precaução (Gooijer e Hyndman, 2006). Supondo-se que os erros são normalmente distribuídos, a variância é dada por

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{2.75}$$

onde α_0, α_1 e β_1 são constantes reais e positivas.

Assim sendo, σ_t^2 segue um modelo GARCH, onde (p,q) representam a ordem do componente GARCH(p,q). A variância é dada da seguinte forma:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i x_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^q \tag{2.76}$$

onde as restrições são dadas por $\alpha_i > 0$, $i = 1,2,\ldots,p$; $\beta_j > 0$, $j = 1,2,\ldots,p$ e $\alpha_i + \beta_j < 1$.

(vii) Modelo Autoregressivo Heterocedástico Condicional Generalizado Exponencial (EGARCH): O modelo EGARCH(p,q) descreve as diferentes respostas da taxa de

retorno aos choques positivos e negativos, mas sem a necessidade de qualquer restrição paramétrica. O modelo é formalizado da seguinte maneira:

$$\ln(\sigma_t^2) = \alpha_0 + \gamma_i x_{t-1}^2 + \beta_j \ln(\sigma_{t-j}^2) \tag{2.77}$$

A assimetria é capturada pelo coeficiente γ. Se $\gamma < 0$, um choque negativo aumenta a volatilidade dos retornos. Se $\gamma > 0$, um choque positivo diminui a volatilidade dos retornos. Se $\gamma = 0$, há ausência de assimetria na volatilidade dos retornos. O coeficiente β indica a persistência de choques na volatilidade.

(viii) Modelo Autoregressivo Heterocedástico Condicional Generalizado de Glosten, Jagannathan e Runkle (GJR – GARCH): O modelo GJR-GARCH$(1,1)$, é apresentado pela equação:

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma d_{t-1} x_{t-1}^2 \tag{2.78}$$

onde d_{t-1} é a variável *dummy*, tal que $d_{t-1} = 1$, se $x_{t-1} < 0$; e $d_{t-1} = 0$, se $x_{t-1} > 0$. Se $\gamma = 0$, não haverá efeito assimétrico. A assimetria é capturada pelo coeficiente γ que indica a influência com que os choques negativos apresentam impactos maiores do que os positivos sobre a volatilidade. O coeficiente β_1 mede a persistência dos choques nas variâncias futuras.

2.5 MODELOS NÃO-LINEARES

O problema de identificação no domínio do tempo para o caso linear ou não-linear é determinar a relação entre as entradas e saídas passadas, as saídas futuras. Se um número finito de entradas $u(t)$ e saídas $y(t)$ passadas forem coletados em um vetor $\wp(t)$, tal que

$$\wp(t) = \left[y(t-1)...y(t-n_y), u(t-1)...u(t-n_u)\right]^T \tag{2.79}$$

onde n_y é o número de saídas atrasadas e n_u é o número de entradas atrasadas, então o problema deve ser compreendido como o relacionamento da função f entre a próxima saída $y(t)$ e $\wp(t)$, tal que

$$y(t) \rightarrow f(\wp(t)) \tag{2.80}$$

Para se obter essa compreensão, é necessário um conjunto de dados observados, o qual consiste em dados de entradas $u(t)$ e saídas $y(t)$ observadas do sistema real para que o vetor $\wp(t)$ possa ser construído. A função f pode ser qualquer função, linear ou não-linear, e definirá a estrutura do modelo (Ljung e Glad, 1994; Aguirre, 2000; Ahmad e Sutton, 2000). Segundo Santos (2005), a partir dessa definição geral, os modelos não-lineares se subdividem em:

- Modelos auto-regressivos não-lineares, envolvendo apenas funções das variáveis dependentes (modelos univariados). Tipicamente, apenas funções matemáticas simples têm sido consideradas (por exemplo: funções seno, co-seno, logística, entre outras);
- Modelos de função de transferência não-lineares, usando funções da variável dependente defasada e funções de variáveis explicativas defasadas e correntes;
- Modelos bi-lineares, dados pela relação $y_t = \sum_{j,k} \beta_{j,k} y_{t-j} e_{t-k}$. Quando acrescidos do produto de uma variável explicativa *x*(*t*) e um termo residual defasado, têm-se os modelos bi-lineares multivariados;
- Modelos médias móveis não-lineares, formados pela soma de funções dos resíduos defasados $e_t, e_{t-1}, ..., e_{t-j}$;
- Modelos duplamente estocásticos, formados pelos produtos cruzados à variável dependente defasada e às variáveis explicativas correntes e defasadas.

Existe uma grande variedade de representações não-lineares, que a princípio, podem ser utilizados na identificação de sistemas e algumas para a previsão de séries temporais. Porém, não é objetivo da seção 2.5 listar um grande número de representações, nem mesmo detalhar qualquer uma delas. Neste contexto, é fornecida uma visão geral de alguns modelos existentes, como os modelos de Volterra e os modelos de Hammerstein e Wiener. Outra representação não-linear são os sistemas nebulosos, que será abordada no capítulo 3.

2.5.1 Modelo de Volterra

A saída $y(t)$ de um sistema não-linear com entrada $u(t)$ pode ser representada pela chamada série de Volterra definida como:

$$y(t) = h_0 + \sum_{j=1}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_j(\tau_1, \dots, \tau_t) \prod_{i=1}^{j} u(t-\tau_i) d\tau_i \tag{2.81}$$

onde as funções h_j são os núcleos ou mesmo generalizações não-lineares da resposta ao impulso $h_1(t)$. De fato, para um sistema linear $j=1$, a equação (2.81) se reduz à integral de convolução.

Esta equação (2.81) representa a soma das saídas dos subsistemas chamados de funcionalidades de Volterra, conforme apresentado na figura 2.1 (Aguirre, 2000; Billings, 1980).

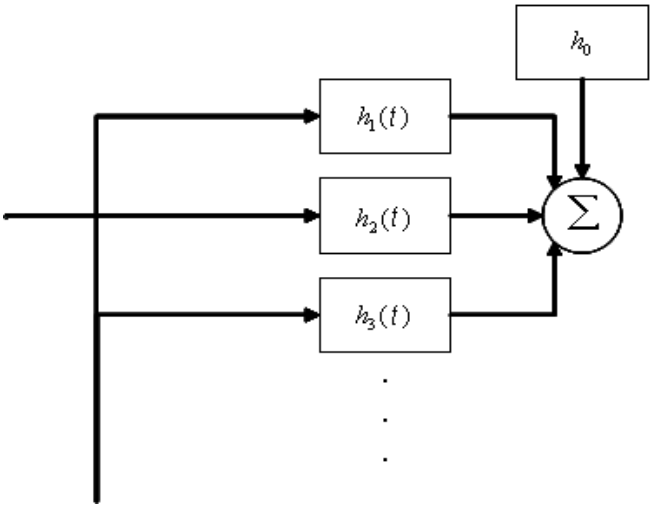

Figura 2.1 Representação gráfica da série de Volterra.

2.5.2 Modelo de Hammerstein e de Wiener

Os modelos de Hammerstein e de Wiener foram dois tipos de modelos não-lineares populares até duas ou três décadas atrás (Nepomuceno, 2002; Greblick e Pawlak, 1989). Ambos são uma composição de um modelo dinâmico linear $H(s)$ em cascata com uma função estática não-linear $f(.)$. No caso do modelo de Hammerstein, a não-linearidade estática precede o modelo dinâmico linear, ou seja:

$$U(s) = f\big(U(s)\big) \text{ e } Y(s) = H(s)Us) \tag{2.82}$$

onde $U(s)$ é a entrada e $Y(s)$ a saída do sistema no domínio da freqüência.

No caso do modelo de Wiener, o modelo dinâmico linear procede a não-linearidade estática, isto é:

$$Y(s) = H(s)U(s) \text{ e } Y(s) = f\big(Y(s)\big) \tag{2.83}$$

Em problemas de identificação, apenas tem-se acesso aos sinais $y(t)$, $u(t)$ e a função $f(.)$ é desconhecida (Aguirre, 2000; Cassini e Aguirre, 1999).

2.6 RESUMO DO CAPÍTULO

Mueller (1996) coloca como objetivo inicial da análise de séries temporais a realização de inferências sobre as propriedades ou características básicas do mecanismo gerador do processo estocástico das observações da série. Assim, através da abstração de regularidades contidas nos fenômenos observáveis de uma série temporal existe a possibilidade de se construir um modelo matemático como uma representação simplificada da realidade (Mueller, 1996).

Após a formulação do modelo matemático, obtido pela seleção entre as alternativas de classes de modelos identificadas como apropriadas para essa representação e subseqüente estimação de seus parâmetros, é possível utilizá-lo para testar alguma hipótese ou teoria e realizar a previsão de valores futuros da série temporal (Mueller, 1996).

Capítulo 3

SISTEMAS NEBULOSOS

O conceito de conjunto nebuloso (*fuzzy set*) foi introduzido por Zadeh (1965) e desde então, os sistemas baseados em conjuntos nebulosos (sistemas nebulosos), vem avançando tanto do ponto de vista teórico quanto prático. Os termos conjunto nebuloso e lógica nebulosa têm sido freqüentemente utilizados como sinônimos, provendo uma base para geração de técnicas para solução de problemas, com uma vasta aplicabilidade nas áreas de controle de processos, tomada de decisão e modelagem.

A lógica nebulosa deriva de sua habilidade em gerar respostas baseadas em informações vagas, ambíguas e verdades parciais, baseando-se na teoria dos conjuntos nebulosos (Sandri e Correia, 1999). Esta é uma generalização da proposição da lógica clássica, onde a informação é “completamente verdadeira” ou “completamente falsa”. Entretanto, na lógica nebulosa o grau de verdade varia entre 0 e 1, o que leva a ser “parcialmente verdadeira” ou “parcialmente falsa”. Isso faz dela uma das tecnologias para o desenvolvimento de sistemas para controlar processos sofisticados e complexos, de forma simples e de baixo custo (Sandri e Correia, 1999). Entretanto, é importante salientar que estas duas nomenclaturas tratam aspectos distintos da teoria.

Lógica Nebulosa: A lógica nebulosa pode ser vista como uma extensão da lógica bi-valorada. Ao invés de considerar apenas dois valores-verdade (Verdadeiro=0 e Falso=1) como é o caso da lógica bi-valorada, ou um intervalo unitário como no caso da lógica multi-valorada, a lógica nebulosa pressupõe que os valores-verdade são conjuntos nebulosos definidos no intervalo [0,1]. A definição de regras nebulosas com base em implicações lógicas permite o uso do termo lógica nebulosa para descrever a computação baseada em regras nebulosas. A teoria associada à lógica nebulosa

requer a definição de conceitos distintos, tais como valor-verdade, proposições e implicações nebulosas.

Conjuntos Nebulosos: A designação de conjuntos nebulosos é geralmente usada para descrever a teoria de conjuntos nebulosos que engloba inclusive alguns conceitos de lógica nebulosa. Vista como uma extensão da teoria de conjuntos clássicos, a teoria de conjuntos nebulosos está associada aos conceitos básicos de funções de pertinência, operações com conjuntos nebulosos, números nebulosos, relações nebulosas, regras nebulosas, regra composicional de inferência, entre outras.

A utilização de conjuntos nebulosos é baseada na regra composicional de inferência, com seus operadores associados, que resulta em sistemas de decisão, estruturados no formato de uma base de regras nebulosas, adequados para a implementação de processos dedutivos. Estes sistemas, conhecidos como sistemas de inferência nebulosa ou simplesmente sistemas nebulosos, dependem da especificação de uma série de elementos, que incluem: a quantidade e o tipo de regras nebulosas, os parâmetros das funções de pertinência, a semântica das regras que participam do raciocínio aproximado e os operadores do mecanismo de inferência utilizado para obter uma saída, a partir dos dados de entrada. Os sistemas baseados em regras nebulosas possuem grande habilidade em lidar com a ambigüidade e a subjetividade presente no raciocínio humano (Pedricz e Gomide, 1998). O uso de sistemas nebulosos é indicado quando:

- O modelo matemático é complexo para ser rapidamente avaliado em tempo real ou requer muita memória para ser implementado fisicamente;
- O processo envolve a interação com um operador humano, ou especialista, preparado para especificar os parâmetros do conjunto de regras a ser utilizado no sistema nebuloso.

Pode-se acrescentar a esses itens a possibilidade de se ter um projeto automático para o ajuste de parâmetros na ausência do especialista e o interesse pela

representação do conhecimento por uma base de regras lingüísticas, criando o cenário ideal à utilização dos sistemas nebulosos.

O formato de uma regra do modelo lingüístico possui uma parte premissa e uma parte conseqüente do tipo:

$$\text{SE } x \text{ é } C \text{ (premissa) ENTÃO } y \text{ é } D \text{ (conseqüente)} \qquad (3.1)$$

onde x e y são as variáveis de entrada e saída, respectivamente, e C e D são termos lingüísticos associados aos conjuntos nebulosos que descrevem lingüisticamente essas variáveis. Para um dado valor de entrada, a saída correspondente é calculada a partir do conjunto de regras através de regras de um método de inferência (Machado, 2003).

Basicamente, um modelo nebuloso providencia uma forma efetiva de explorar e apresentar imprecisões e aproximações do mundo real. Em particular, um modelo nebuloso parece ser útil quando um sistema não é apropriado para análise por técnicas quantitativas convencionais, ou quando a informação disponível no sistema é incerta ou imprecisa (Xu e Khoshgoftaar, 2004).

Procedimentos de identificação nebulosa usualmente citam técnicas e algoritmos para construção de modelos nebulosos a partir de dados. A construção de um modelo nebuloso de Takagi-Sugeno a partir dos dados medidos do processo (ou sistema) a ser identificado envolve duas tarefas, a identificação da estrutura e a identificação de parâmetros (Coelho e Alves, 2006; Xu e Khoshgoftaar, 2004). As duas etapas importantes da identificação da estrutura são: (i) a determinação do número de regras **SE-ENTÃO** e (ii) a partição do espaço de entrada para um dado conjunto de conjuntos nebulosos com funções de pertinência. A identificação de parâmetros envolve a identificação dos parâmetros das funções de pertinência e os parâmetros dos conseqüentes funcionais (coeficientes de equações lineares).

Alguns estudos recentes destes sistemas apontam para aplicações em áreas diversificadas, tais como controle de processos, reconhecimento de padrões, sistemas de visão computacional e métodos de vida artificial. Um número significativo de implementações vem consolidando os sistemas nebulosos não só em instalações

industriais, mas também, em muitos produtos manufaturados de uso diário (Pedrycz e Gomide, 1998) e até em sistemas de informação geográfica – GIS (Burlamaqui e Liang, 2002).

É apresentado nas próximas seções o conceito fundamental para o entendimento e utilização dos sistemas nebulosos.

3.1 CONJUNTOS NEBULOSOS (*FUZZY SETS*)

A noção de conjunto ocorre freqüentemente quando se tenta organizar, resumir e generalizar o conhecimento a respeito de objetos. Seja ***X*** uma coleção de objetos denominados genericamente por x, então, um conjunto ***A*** é definido por uma coleção de pares ordenados, tal que

$$A = \{(x, \mu_A(x)) | x \in X\} \tag{3.2}$$

A função $\mu_A(x)$ é denominada função de pertinência e determina com que grau um objeto x pertence a um conjunto ***A***, sendo ***X*** denominado de universo de discurso. Em conjuntos clássicos, apenas dois valores para $\mu_A(x)$ são permitidos: o elemento pertence ou não pertence a um determinado conjunto. Na teoria dos conjuntos nebulosos a transição entre pertencer e não pertencer é gradual.

Como exemplo, considere ***X*** = {1, 2, 3, 4, 5, 6, 7, 8, 9} uma coleção de números inteiros. Seja ***A*** um conjunto nebuloso que define "números inteiros próximos a 5", dado por *A* = {(1, 0), (2, 0.4), (3, 0.6), (4, 0.8), (5, 1), (6, 0.8), (7, 0.6), (8, 0.4), (9, 0)}.

Então, o conjunto nebuloso ***A*** pode ser definido como uma coleção de objetos com valores de pertinência variando entre 0 (exclusão completa) e 1 (pertinência completa). Os conjuntos nebulosos representam, portanto, uma generalização dos conjuntos clássicos. O conceito de função de pertinência, este fundamental para a teoria dos conjuntos nebulosos, será discutido na seção a seguir.

3.1.1 Funções de Pertinência

Toda a teoria dos conjuntos está baseada no conceito de pertinência. Um conjunto nebuloso é definido pela função de pertinência $\mu_A(x)$ que estabelece para cada x um grau de pertinência ao conjunto A, com $\mu_A(x) \in [0,1]$. Os conjuntos clássicos podem ser vistos como um caso particular dos conjuntos nebulosos, no qual apenas os limites do intervalo são utilizados na definição da função de pertinência: $\mu_A(x) \in [0,1], \quad x \in X$.

Por exemplo, considerando-se o universo contínuo $x \in \Re$, o conceito de números próximos a "5" pode ser expresso de forma diferente dependendo da definição da função de pertinência associada. Na figura 3.1 é apresentada a diferença entre as funções de pertinência no caso de conjuntos clássicos e conjuntos nebulosos.

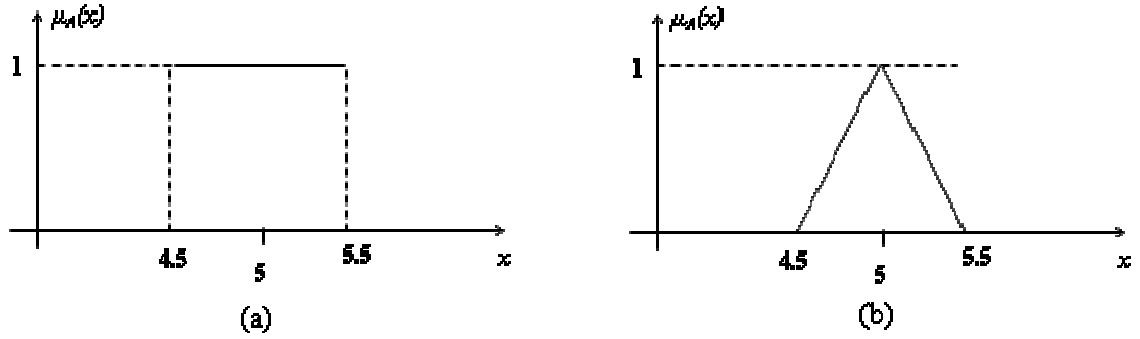

Figura 3.1: Ilustração das diferenças entre as funções de pertinência no caso de conjuntos clássicos (a) e conjuntos nebulosos (b).

- Números próximos a "5" na concepção clássica, onde $\mu_A(x) = 0$ se $4{,}5 \geq x$ ou $x > 5{,}5$ e $\mu_A(x) = 1$ se $4{,}5 < x \leq 5{,}5$.

- Números próximos a "5" na concepção de conjuntos nebulosos, tal que $\mu_A(x) = 0$ se $x \leq 4{,}5$, $\mu_A(x) = \dfrac{x - 4{,}5}{0{,}5}$ se $4{,}5 < x \leq 5$, $\mu_A(x) = \dfrac{5{,}5 - x}{0{,}5}$ se $5 < x \leq 5{,}5$ e $\mu_A(x) = 0$ se $x > 5{,}5$. Em geral, o formato das funções de pertinência é restrito a uma certa classe de funções, representadas por alguns parâmetros específicos (figura 3.2). Os formatos

mais comuns são: triangular, trapezoidal e Gaussiana. Além dos formatos tradicionais, existe uma forma utilizada com freqüência em aplicações práticas: o conjunto unitário (*singleton*).

• Função Triangular: possui os parâmetros (a,m,b) com $a \le m \le b$, tal que $\mu_A(x) = 0$ se $x \le a$, $\mu_A(x) = \frac{x-a}{m-a}$ se $a < x \le m$, $\mu_A(x) = \frac{b-x}{b-m}$ se $m < x \le b$ e $\mu_A(x) = 0$ se $x > b$. Na figura 3.2(a), tem-se o número nebuloso triangular representado por a, m e b. Os parâmetros a e b representam o limite inferior e superior da área de avaliação e o parâmetro m apresenta o grau máximo de pertinência de $\mu_A(x)$, ou seja $\mu_A(x) = 1$.

• Função Trapezoidal: é regida pelos parâmetros (a,m,n,b) com $a \le m$; $n \le b$; $m < n$, tal que $\mu_A(x) = 0$ se $b < x \le a$, $\mu_A(x) = \frac{x-a}{m-a}$ se $a < x \le m$, $\mu_A(x) = \frac{b-x}{b-n}$ se $n < x \le b$ e $\mu_A(x) = 1$ se $m < x \le n$. Na figura 3.2(b), tem-se o número nebuloso triangular representado por a, m, n e b. Os parâmetros a e m representam o segmento ascendente do trapézio e o segmento descendente é dado por n e b.

• Função Gaussiana: possui parâmetros (m,σ_k) com $\sigma_k > 0$, a função Gaussiana é dada por $\mu_A(x) = \exp^{-\sigma_k (x-m)^2}$. Na figura 3.2(c), tem-se o valor modal dado por m e a dispersão por σ_k.

• Conjunto Unitário (*singleton*): têm parâmetros (m, h), tal que $\mu_A(x) = h$ se $x = m$, caso contrário $\mu_A(x) = 0$. Na figura 3.2(d), tem-se o número nebuloso representado por m. O parâmetro m apresenta o grau máximo de pertinência de $\mu_A(x)$, ou seja $\mu_A(x) = h$.

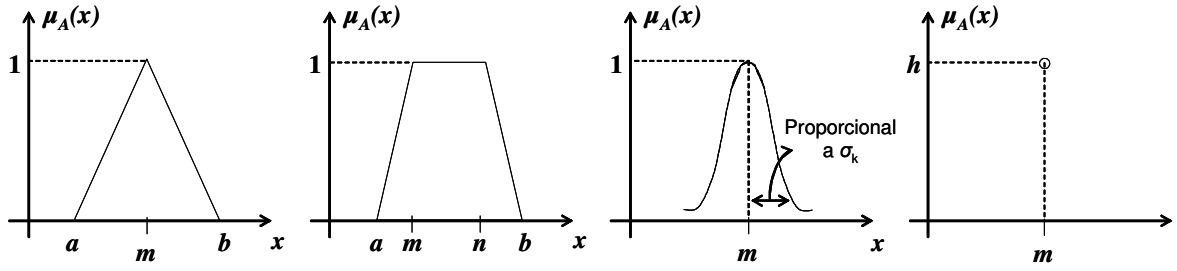

Figura 3.2: Ilustração das diferenças entre as funções de pertinência, especificando os parâmetros associados às funções triangular (a), trapezoidal (b), Gaussiana (c) e conjunto unitário (d).

A escolha do formato adequado nem sempre é óbvia, podendo inclusive não estar ao alcance do conhecimento de um especialista para a aplicação em questão. No entanto, existem os sistemas nebulosos cujos parâmetros das funções de pertinência são definidos pelo especialista. Nestes casos, a escolha de funções triangulares e trapezoidais é comum porque a idéia de se definirem regiões de pertinência total, média e nula é mais intuitiva do que a especificação do valor modal (m) e dispersão (σ_k), conceitos estes ligados às funções Gaussianas. Entretanto, existe cada vez mais uma tendência de sistemas nebulosos adaptativos, nos quais, os parâmetros das funções de pertinência são ajustados, no sentido de otimizar algum objetivo definido a partir dos dados de entrada e saída. Por exemplo, a função Gaussiana é utilizada em sistemas que ajustam estes parâmetros pelo método do gradiente, devido as suas propriedades matemáticas de continuidade e diferenciabilidade.

Outro tipo de função de pertinência associada a uma variável qualquer x será utilizado para indicar a irrelevância desta variável na regra nebulosa. Esta condição irrelevante será representada (em proposições que utilizam $\in$ como agregação) por uma função denominada função de irrelevância, definida por

$$\text{Função de Irrelevância: } \mu_A(x) = 1 \quad \forall x \in \mathrm{X}. \tag{3.3}$$

3.1.2 Definições Básicas em Conjuntos Nebulosos

Embora a função de pertinência, especificada pelos parâmetros que define o seu formato, seja usualmente utilizada para representar um conjunto nebuloso, existem outros parâmetros que podem ser usados para caracterizar os conjuntos nebulosos:

- Suporte (S_A): conjunto dos elementos do universo, para os quais o grau de pertinência é maior do que zero, tal que

$$S_A = \{x | \mu_A(x) > 0\}. \tag{3.4}$$

- Núcleo (*core* - N_A): conjunto dos elementos do universo com grau de pertinência igual a 1, onde

$$N_A = \{x | \mu_A(x) = 1\}. \tag{3.5}$$

- Altura (H_A): valor máximo da função de pertinência, dado por:

$$H_A = \sup_x \{\mu_A(x)\}. \tag{3.6}$$

- *α-corte:* é o conjunto dos elementos do universo para os quais os graus de pertinência são superiores ou iguais a α, onde

$$C_{\alpha A} = \{x | \mu_A(x) \geq \alpha\}. \tag{3.7}$$

Uma outra forma de representação é via *α-cortes*, ilustrada na figura 3.3. Esta forma permite representar um conjunto nebuloso como uma família de outros conjuntos nebulosos particulares e determina a base do teorema da representação (Pedrycz e Gomide, 1998). Outras definições importantes são ilustradas na figura 3.4.

• Conjunto Normal: um conjunto nebuloso é dito normal se sua altura é igual a 1, ou seja, $H_A = 1$. Caso contrário é chamado de subnormal. O núcleo de um conjunto subnormal é um conjunto vazio.

• Conjunto Convexo: um conjunto nebuloso é convexo se sua função de pertinência é dada por

$$\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq \min[\mu_A(x_1), \mu_A(x_2)] \tag{3.8}$$

para todo $x_1, x_2 \in X$ e $\lambda \in [0, 1]$.

• Subconjunto Nebuloso: se $\mu_A(x) \leq \mu_B(x)$, para todo x então $A \subseteq B$, ou seja, A é subconjunto de B. Uma propriedade derivada da equação anterior determina que, se $A \subseteq B \, e \, B \subseteq C \rightarrow A \subseteq C$.

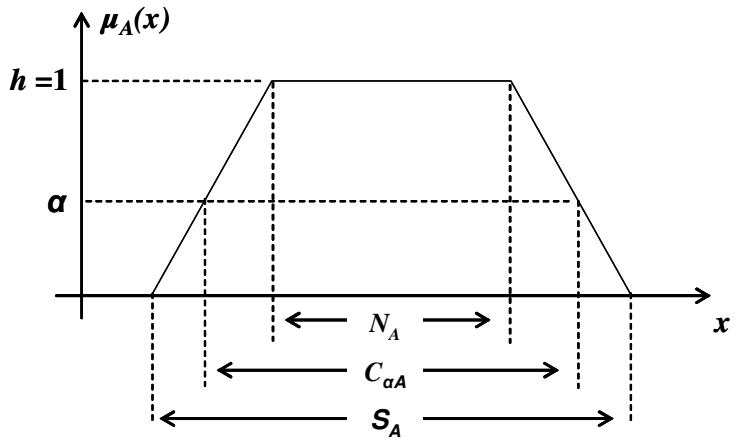

Figura 3.3: Ilustração do suporte (SA), α-corte (CαA), núcleo (NA) e altura (HA).

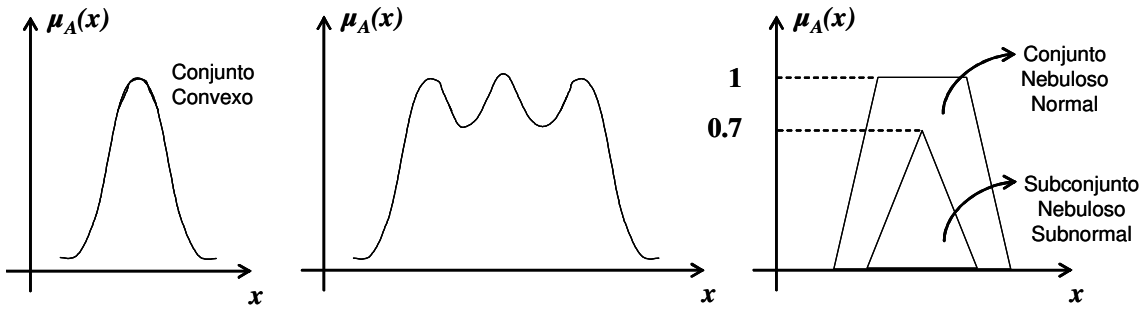

Figura 3.4: Ilustração de convexidade e normalidade em conjuntos nebulosos.

3.1.3 Operações com Conjuntos Nebulosos

Existem várias operações que podem ser aplicadas a conjuntos nebulosos. Há operações com argumento único, que modificam o formato da função de pertinência. Dentre as mais comuns pode-se citar: (i) normalização, converte um conjunto subnormal em um normal; (ii) concentração, que diminui os valores da função de pertinência; (iii) dilatação, que aumenta os valores da função de pertinência; e (iv) intensificação de contraste, que dilata o conjunto para valores de pertinência acima de 0,5 concentrando o conjunto para valores abaixo de 0,5 (Pedrycz e Gomide, 1998). Há operações de múltiplos argumentos, que envolvem a combinação, agregação e comparação de dois ou mais conjuntos nebulosos. A maioria destas operações é derivada da teoria de conjuntos clássicos. A seguir, serão descritas a operação de complemento e as operações de múltiplos argumentos mais comumente aplicadas a conjuntos nebulosos.

• União, Intersecção e Complemento: são operações essenciais realizadas em conjuntos clássicos, ilustrada na figura 3.5. A tabela 3.1 resume algumas propriedades básicas destas operações, considerando os conjuntos clássicos A e B definidos num universo X. Com base na teoria dos conjuntos clássicos, essas operações para conjuntos nebulosos, foram definidas a partir da função de pertinência:

União: $\mu_{A\ (A\cup B)}(x) = \max[\mu_A(x), \mu_B(x)] = \mu_A(x) \vee \mu_B(x)$ (3.9)

Intersecção: $\mu_{A\ (A\cap B)}(x) = \min[\mu_A(x), \mu_B(x)] = \mu_A(x) \wedge \mu_B(x)$ (3.10)

Complemento: $\bar{\mu}_A(x) = 1 - \mu_{A(A)}(x)$ (3.11)

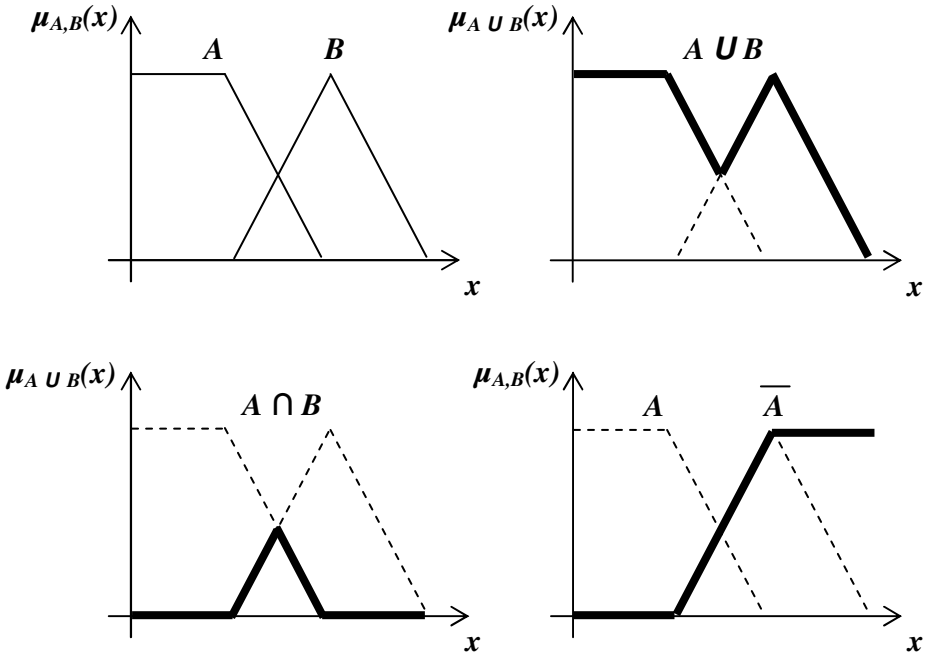

Figura 3.5: Ilustração de operações com os conjuntos A e B (a). União (b), Intersecção (c) e Complemento (d).

Devido ao isomorfismo entre a teoria dos conjuntos e a lógica proposicional bi-valores, a intersecção e a união podem ser identificados pela conjunção (***AND***) e pela disjunção (***OR***), respectivamente. Assim, as operações de conjunção e disjunção podem ser representadas pelos operadores $\wedge$ e $\vee$ (Pedrycz e Gomide, 1998).

• Normas e Co-normas Triangulares: formam uma classe geral de operadores de união e intersecção, com características de comutatividade, associatividade e monotonicidade, atendendo ainda as condições de contorno, conforme apresentado na tabela 3.1. Diferente da união e intersecção, que trabalham com conjuntos definidos

num mesmo universo, as operações baseadas em normas e co-normas triangulares pode operar conjuntos em universos distintos.

Tabela 3.1: Propriedades das operações com conjuntos clássicos.

Comutatividade	$A \cup B = B \cup A$, $A \cap B = B \cap A$
Associatividade	$(A \cup B) \cup C = A \cup (B \cup C)$ $(A \cap B) \cap C = A \cap (B \cap C)$
Distributividade	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
Idempotência	$A \cup A = A$, $A \cap A = A$
Condições Limites	$A \cup \phi = A$, $A \cup X = X$ $A \cap \phi = \phi$, $A \cap X = A$
Involução	$\overline{\overline{A}} = A$
Lei da Contradição	$A \cap \overline{A} = \phi$
Lei do Meio Excluído	$A \cup \overline{A} = X$
Lei de Morgan	$\overline{A \cup B} = \overline{A} \cap \overline{B}$ $\overline{A \cap B} = \overline{A} \cup \overline{B}$

Sejam A e B dois conjuntos nebulosos definidos nos universos X e Y, respectivamente, e a e b valores de pertinência dados por $\alpha = \mu_A(x)$ e $\mu_B(y)$. Então, as normas e co-normas triangulares (norma-t e norma-s) podem ser definidas como:

i. Norma-*t*: operador de dois argumentos $t: [0,1]^2 \rightarrow [0,1]$ que satisfaz as seguintes condições:

Comutatividade: $a \; t \; b = b \; t \; a$;

Associatividade: $a \; t \; (b \; t \; c) = (a \; t \; b) \; t \; c$;

Monotonicidade: se $a \leq b$ e $c \leq d$, então $a \; t \; c \leq b \; t \; d$;

Condições de Contorno: $0 \; t \; a = 0$, $1 \; t \; a = a$.

ii. Norma-*s*: operador de dois argumentos $s: [0,1]^2 \rightarrow [0,1]$ que satisfaz as seguintes condições:

Comutatividade: $a \; s \; b = b \; s \; a$;

Associatividade: $a \; s \; (b \; s \; c) = (a \; s \; b) \; s \; c$;

Monotonicidade: se $a \leq b$ e $c \leq d$, então $a \; s \; c \leq b \; s \; d$;

Condições de Contorno: $0 \; s \; a = 0$, $1 \; s \; a = a$.

Em geral, as normas t e s não satisfazem, necessariamente, as demais propriedades válidas para os conjuntos clássicos. Por exemplo, a lei da contradição e o princípio da exclusão do meio são características presentes apenas nas normas t_2 e s_2, considerando-se $p_t = 0$. Para as outras normas listadas anteriormente, essas duas condições não se aplicam. Para exemplificar, considere a norma-$t = t_9$ = ***MIN*** e a co-norma-$s = s_9$ = ***MAX***, e um valor de pertinência $\alpha = \mu_A(x) = 0{,}5$. Então:

$$A \cap \overline{A} = a \, \mathrm{t}_9 \, (1 - a) = \min [a, 1 - a] = 0{,}5 \neq \phi \tag{3.12}$$

$$A \cup \overline{A} = a \, \mathrm{s}_9 \, (1 - a) = \max [a, 1 - a] = 0{,}5 \neq X \tag{3.13}$$

O princípio da idempotência só se aplica às normas ***MIN*** e ***MAX***.

$$a \, \mathrm{t}_9 \, a = \min [a, a] = a \tag{3.14}$$

$$a \, \mathrm{s}_9 \, a = \max [a, a] = a \tag{3.15}$$

Para outras normas, apenas pode-se garantir que:

$$\underbrace{atatata\cdots ata}_{n\ vezes} \le \underbrace{atata\cdots ata}_{n-1\ vezes}; \tag{3.16}$$

$$\underbrace{asasasa\cdots asa}_{n\ vezes} \ge \underbrace{asasa\cdots asa}_{n-1\ vezes}; \tag{3.17}$$

Assim como a idempotência, a propriedade de distributividade está presente apenas nas normas ***MIN*** e ***MAX***. As normas t e s são consideradas operadores de agregação, pois podem ser usadas para combinar uma coleção de conjuntos nebulosos para produzir um único conjunto, também nebuloso.

3.1.4 Sistemas Nebulosos

Os sistemas nebulosos, também conhecidos como sistemas de inferência nebulosa, sistemas nebulosos baseados em regras ou modelos nebulosos representam a mais importante ferramenta de modelagem baseada na teoria dos conjuntos nebulosos.

Os sistemas nebulosos têm sido aplicados com sucesso em diversas áreas, tais como: controle automático (Takatsu e Itoh, 1999), tomada de decisão (Simões, 2003), classificação e reconhecimento de padrões (Bonventi e Costa, 2003), sistemas inteligentes, previsão de séries temporais e robótica (Dote e Ovaska, 2001), entre outros.

Um sistema de inferência nebulosa é um mapeamento ou função de um espaço de alternativas de entrada para um espaço de saída. A figura 3.6 descreve a estrutura básica de um sistema nebuloso puro, que possui três componentes conceituais: uma base de regras, que contém o conjunto de regras nebulosas (regras de produção); uma base de dados, que define as funções de pertinência usadas nas regras nebulosas; e um mecanismo de inferência, que realiza um procedimento de inferência (raciocínio nebuloso) para obter a saída ou conclusão, baseado em regras e fatos conhecidos.

Uma característica do sistema nebuloso puro é que suas entradas e saídas são conjuntos nebulosos, que geralmente são regidos por termos de linguagem natural.

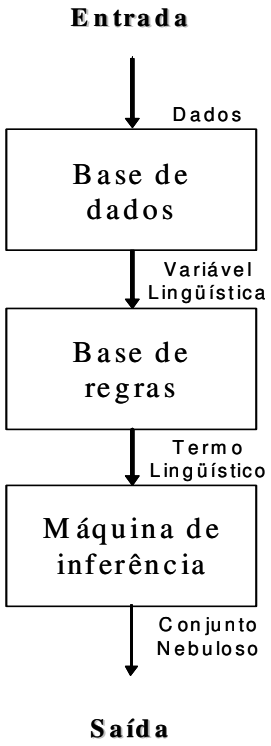

Figura 3.6: Ilustração da estrutura básica de um sistema nebuloso puro.

Em Abonyi (2003), Renners (2004) e Sandri e Correia (1999), o mecanismo de raciocínio é desdobrado em unidade de nebulização, máquina de inferência e unidade de desnebulização. Em Babuška (1998), foram anexados filtros dinâmicos aos sinais de entrada e saída para proporcionar comportamentos dinâmicos ao sistema (Babuška, 1998), resultando no modelo descrito na figura 3.7.

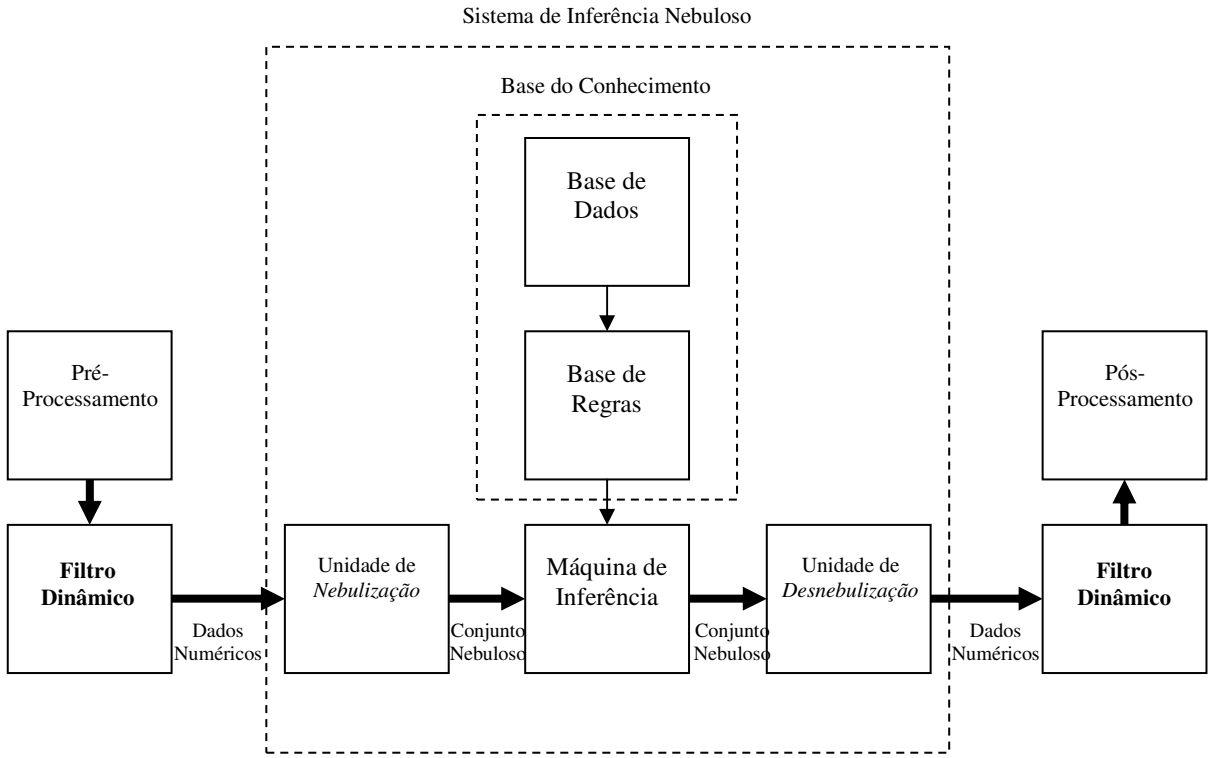

Figura 3.7: Sistema de inferência nebuloso com nebulizador e desnebulizador. O fluxo da informação é indicado pela linha mais fina e o fluxo computacional é indicado pela linha mais grossa.

Com relação à base de conhecimento, a mesma é constituída pela base de dados e pela base de regras de maneira a caracterizar o funcionamento completo do sistema de inferência. Na base de dados estão armazenadas as variáveis lingüísticas, as definições dos respectivos universos de discurso e as funções de pertinência caracterizando os termos lingüísticos utilizados para cada variável lingüística. Na base de regras estão as declarações lingüísticas do tipo **SE–ENTÃO**, definidas por um especialista ou extraídas de dados numéricos (como por exemplo, em problemas de previsão de séries temporais) que constituem aspecto fundamental ao desempenho do sistema de inferência nebulosa. A unidade de nebulização ou *fuzificador* mapeia os números precisos da entrada em conjuntos nebulosos para ativação das regras relevantes em uma dada aplicação. A máquina de inferência mapeia os conjuntos nebulosos da entrada em conjuntos nebulosos na saída, de acordo com as características da base de dados e da base de regras, combinando as regras que foram

ativadas. A unidade de desnebulização ou *defuzificador* mapeia os conjuntos nebulosos na saída em números precisos, os quais podem corresponder a ações de controle em um sistema dinâmico ou à predição da sua saída num instante de tempo futuro. As relações entre as variáveis são representadas por meio de regras com a seguinte forma geral: **SE** antecedente **ENTÃO** conseqüente. O antecedente consiste em uma proposição nebulosa do tipo $\tilde{n}$ *é* F, onde $\tilde{n}$ é uma variável lingüística e F é um conjunto nebuloso definido como termo lingüístico. Dependendo da forma do conseqüente, existem alguns tipos de sistemas de inferência nebulosos, como os descritos na seção 3.1.6.

A principal característica do sistema nebuloso com Nebulizador e Desnebulizador é permitir que o usuário entre com os valores reais e obtenha valores reais na saída, mas todo o processamento da informação é realizado em termos lingüísticos (Machado, 2003).

3.1.5 Métodos de Transformação da Saída

Em muitas situações práticas, é necessário que se aplique uma transformação na saída inferida pelo sistema nebuloso. Em problemas de controle de processos, por exemplo, existe a necessidade de uma saída não-nebulosa (*crisp*).

Tal situação pode ser ilustrada pelo sistema de inferência descrito na figura 3.7, onde, uma vez que as regras tenham sido estabelecidas, o sistema nebuloso mapeia as entradas *crisp* – oriundas de um conjunto de dados resultante de medições ou observações experimentais – em saídas *crisp*, tal que

$$y = f(\mathrm{u}) \tag{3.18}$$

onde u e y são a entrada e a saída do sistema de inferência nebuloso e f corresponde à representação quantitativa deste mapeamento.

Em modelos que utilizam conseqüentes lingüísticos, como é o caso dos modelos Mamdani, a obtenção de uma saída *crisp* requer um estágio de transformação da informação nebulosa em informação não-nebulosa (desnebulização). Existem diferentes métodos que implementam esta transformação. Dentre os mais utilizados destacam-se: média dos máximos, centro de massa e centro de área. Seja C um conjunto nebuloso definido no universo Z, então:

• Centro de Massa (CoG): o resultado da transformação de C em um valor não-nebuloso é o centro de massa, dado por:

$$\hat{z} = \frac{\int_z \mu_c(z)\, z\, \partial z}{\int_z \mu_c(z)\, \partial z} \tag{3.19}$$

• Centro de Área (CoA): neste caso, $\hat{z}$ resulta do balanço entre duas áreas de C determinada por $\hat{z}$:

$$\int_{-\infty}^{\hat{z}} \mu_c(z)\partial z = \int_{\hat{z}}^{\infty} \mu_c(z)\partial z \tag{3.20}$$

• Média dos Máximos (MoM): os valores do domínio correspondentes ao máximo da função de pertinência do conjunto C são identificados e a média define o valor não-nebuloso;

Em alguns modelos a saída obtida de cada regra é não-nebulosa (*crisp*), mas para aumentar a interpretabilidade do modelo, a aplicação poderia exigir uma informação lingüística no conseqüente de cada regra. Neste sentido, algumas abordagens têm proposto métodos para a transformação de informação não-nebulosa (*crisp*) em informação nebulosa, representada por funções de pertinência associadas aos termos lingüísticos específicos.

3.1.6 Modelos de Sistemas de Inferência Nebulosas

Existem vários modelos de sistemas de inferência nebulosos que utilizam métodos de inferência. Na maioria dos casos, o antecedente é formado por proposições lingüísticas e a distinção entre os modelos se dá no conseqüente das regras nebulosas. Entre os modelos mais conhecidos pode-se destacar (Delgado, 2002; Machado *et al.*, 2006):

• Sistema de inferência nebulosa de Mamdani: tanto o antecedente como o conseqüente são proposições nebulosas (Mamdani e Assilian, 1975). Em suma, esse modelo utiliza conjuntos nebulosos também nos conseqüentes das regras nebulosas (René, 1995; Castanho, 2005). A saída final é representada por um conjunto nebuloso resultante da agregação da saída inferida de cada regra. Para se obter uma saída final não nebulosa, adota-se um dos métodos de transformação da saída nebulosa em não-nebulosa.

• Sistema de inferência nebulosa de Larsen: deriva do modelo de Mamdami (Machado *et al.*, 2006). Ao invés de utilizar-se do operador de Mamdami nas implicações das regras nebulosas, é utilizado o operador produto. Dessa forma, o modelo nebuloso de Mamdami pode ter variações se forem feitas diferentes escolhas para o conectivo ***AND*** (norma-t) e ***OR*** (co-norma-t).

• Sistema de inferência nebulosa de Takagi-Sugeno: onde o antecedente é uma proposição nebulosa e o conseqüente consiste em uma expressão funcional das variáveis lingüísticas definidas no antecedente. A saída final é obtida pela média ponderada das saídas inferidas de cada regra (Takagi e Sugeno, 1985; Reyes, 2002). Os coeficientes da ponderação são dados pelos graus de ativação das respectivas regras (Abonyi *et al.*, 2002).

• Sistema de inferência nebulosa de Tsukamoto: onde utiliza funções de pertinência no conseqüente. Assim como no modelo Takagi-Sugeno, é inferido um valor não-nebuloso induzido pelo nível de ativação da regra. A saída final do sistema é obtida por média ponderada das saídas inferidas de cada regra (Machado *et al.*, 2006).

A seguir, é detalhado um dos modelos mais encontrados na revisão bibliográfica (Babuška, 1998; René, 1995; Sugeno, 1999; Gorp, 2000; Viljamaa, 2000), o modelo de Takagi-Sugeno.

3.1.6.1 Modelo de Takagi-Sugeno (TS)

Os modelos Takagi-Sugeno recentemente têm-se transformado numa ferramenta eficiente de Engenharia para modelagem, controle, identificação de sistemas e previsão de séries temporais.

Os modelos TS formam uma transição natural entre o controle convencional e o controle baseado em regras, pela expansão e generalização do conceito do bom conhecimento de ganho de programação, utilizando a idéia de linearização em uma região nebulosa definida no estado de espaço e a geração de regras nebulosas a partir de dados de entrada-saída. Devido as regiões nebulosas, o sistema não-linear é decomposto em uma estrutura multi-modelo que consiste em modelos lineares que não são necessariamente independentes.

A representação de modelos TS oferece soluções computacionais eficientes a uma larga escala na introdução de controle de processos em uma estrutura de múltiplos modelos, que são capazes de aproximar a dinâmica não-linear, modelos de operação múltiplos, parâmetros significativos e estruturas variadas (Angelov e Filev, 2004). A base de regras nebulosas típicas no modelo TS é da forma **SE-ENTÃO** (*IF-THEN*) e representada pela figura 3.8. Neste caso, as regras propostas por Takagi-Sugeno (1985) são dadas por (Renners, 2004):

$$\begin{aligned} &\mathrm{R_i}: \mathrm{SE}\ x_1\ \acute{e}\ A_i^1\ ,\ x_2\ \acute{e}\ A_i^2\ ,\ldots,\ x_n\ \acute{e}\ A_i^n \\ &\mathrm{ENTÃO}\ \hat{y}_1 = a_i(x) + b_i\ ,\quad i = 1,2,\ldots,k \end{aligned} \tag{3.21}$$

O antecedente **SE** define a parte antecedente, enquanto as funções da regra **ENTÃO** se constituem na parte conseqüente do modelo nebuloso. A variável k denota

o número de regras na base de regras, R_i é a *i*-ésima regra, $x = [x_1, \cdots, x_n]^T \in \chi$ é o vetor das variáveis de entrada das regras (antecedentes), $A_i^1, \cdots, A_i^n$ são conjuntos nebulosos definidos no espaço dos antecedentes das regras e y_i é a saída da regra. Portanto, a resposta do modelo TS é uma soma ponderada dos conseqüentes, isto é, os parâmetros $a(x), b(x)$ são combinações lineares convexas dos parâmetros conseqüentes (Babuška, 1998).

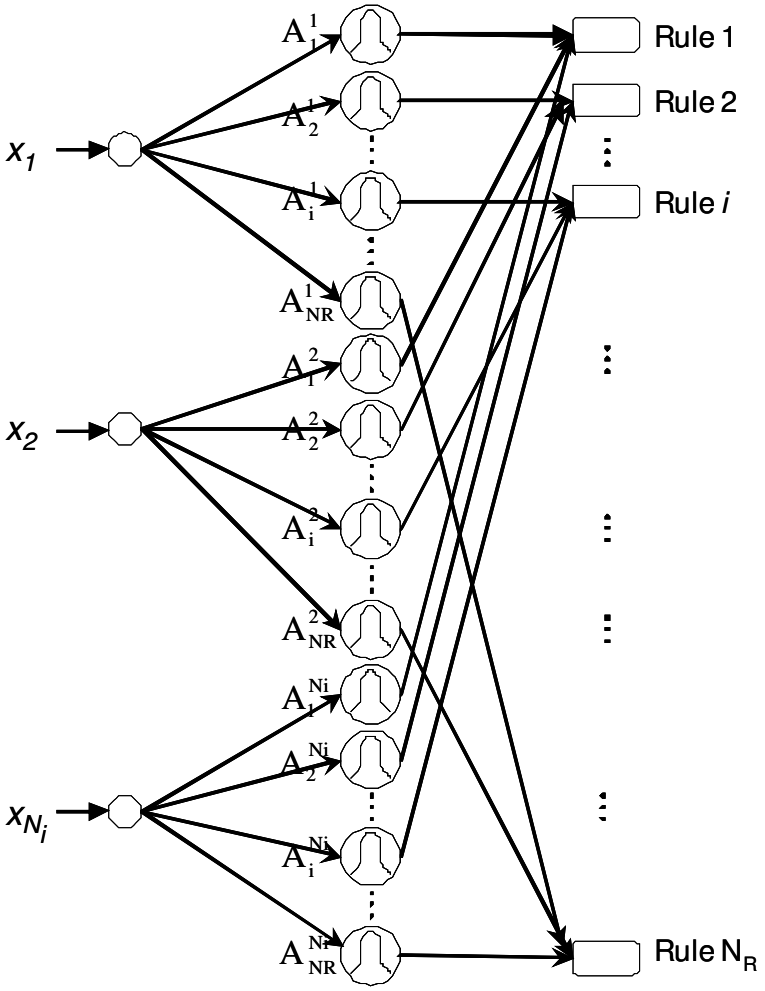

Figura 3.8: Modelo de Takagi-Sugeno.

O algoritmo original de identificação sugerido por Sugeno e seus colaboradores têm as seguintes etapas (Machado, 2003):

(i) Escolher a estrutura da premissa e a estrutura da parte conseqüente;
(ii) Estimar os parâmetros da estrutura determinada no passo (i);
(iii) Avaliar o modelo;
(iv) Repetir a etapa (i) até que o resultado obtido seja satisfatório.

Entretanto, a implementação de tal algoritmo não é trivial, porque para determinar as variáveis das funções de pertinência é necessário resolver um problema de programação não linear (Machado, 2003).

A saída agregada do modelo $\hat{y} \in \Im$ é calculada pela média ponderada dos conseqüentes da regra:

$$\hat{y} = \frac{\sum_{i=1}^{K} \beta_i(x)\, \hat{y}_i}{\sum_{i=1}^{K} \beta_i(x)} \tag{3.22}$$

O grau de ativação normalizado para a regra $\beta_i(x)$ é definido como:

$$\beta_i(x) = \prod_{j=1}^{n} \mu_{Aij}(x_j), \quad i = 1,2,\ldots,k \tag{3.23}$$

onde $\mu_{Aij}(x_j) : \Re \to [0,1]$ é uma função de pertinência do conjunto nebuloso A_{ij} no antecedente de R_i. A cada conjunto nebuloso do antecedente A_{ij} é associada uma função de pertinência $\mu_{Aij}(x_i)$ descrita por

$$\mu_{Aij}(x_i) = \exp\left[-\frac{1}{2} \frac{(x_i - m_{ij})^2}{\sigma_{ij}^2} \right] \tag{3.24}$$

A ordem do polinômio define a ordem do modelo. Os mais comuns são os modelos TS de ordem zero (conseqüentes constantes) e 1ª ordem (conseqüentes lineares). Em

relação ao modelo TS de ordem zero, tem-se aplicação na indústria para interpretação de problemas, devido a estimação simultânea das regras do conseqüente e uma única otimização de mínimos quadrados. Este passo de estimação é referido por Renners (2004) como estimação global.

Quando os pares de dados desejados de entrada-saída são derivados ou adquiridos de uma função desconhecida ou de um dado sistema, o problema da modelagem nebulosa é de construir um modelo nebuloso que possa aproximar corretamente o relacionamento de entrada-saída. Assim, o projeto de um modelo nebuloso consiste em encontrar uma estrutura otimizada (número de regras e parâmetros de entradas) e a otimização dos parâmetros de uma função, de modo a minimizar o erro entre a saída do modelo nebuloso e a saída desejada (Du e Zhang, 2007).

Nesse contexto, um modelo TS tanto pode de ser considerado como um modelo linear variante nos parâmetros (LPV, *Linear Parameter Varying*),como um mapeamento do espaço do antecedente (entrada) à região convexa (politopo) no espaço dos submodelos locais definidos pelos parâmetros do conseqüente, conforme ilustrado na figura 3.9.

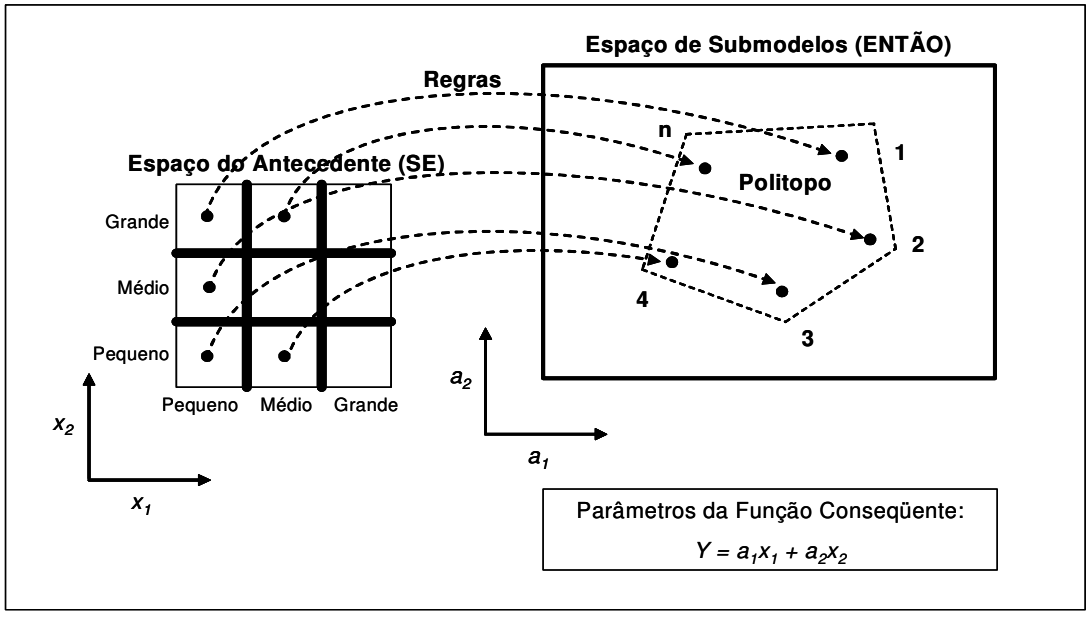

Figura 3.9: Mapeamento nebuloso do espaço de entrada no espaço de submodelos locais (antecedentes). O espaço de entrada do sistema dinâmico é dividido em regiões nebulosas nas quais os submodelos representam as expressões funcionais do conseqüente.

3.2 RESUMO DO CAPÍTULO

Neste capítulo, foram apresentados os fundamentos dos conjuntos nebulosos e também dos sistemas nebulosos.

Conforme já mencionado, existem modelos de sistemas de inferência nebulosos que utilizam métodos de inferência. O procedimento de raciocínio nebuloso é paralelo, ou seja, a combinação entre a parte antecedente da regra com a entrada pode ser calculada de forma paralela. Na maioria dos casos, o antecedente é formado por proposições lingüísticas e a distinção entre os modelos se dá no conseqüente das regras nebulosas.

Entre os modelos mais conhecidos, abordou-se o modelo de Takagi-Sugeno, onde suas vantagens são sua eficiência computacional, que garante a continuidade da superfície de saída, e ser apto à análise matemática, fornecendo promissores resultados quando aplicado a sistemas presentes em Engenharia.

Capítulo 4

AGRUPAMENTO DE DADOS

Este capítulo aborda os conceitos de agrupamento de dados e as suas categorias. Além disso, detalha os principais algoritmos de agrupamento nebuloso de dados existentes na literatura, inclusive nas áreas de identificação de sistemas e previsão, refletindo a importância de explorar e analisar dados.

Um aspecto importante a ser definido pelas técnicas de agrupamentos é o critério de similaridade e dissimilaridade a ser adotado (Valdes, 2004). Dessa forma, pode-se afirmar que quanto maior for a medida de similaridade, maior será a relação entre os dados, assim como, quanto maior for a medida de dissimilaridade, menor será a semelhança entre os dados. Para proporcionar meios de se quantificar a similaridade e dissimilaridade entre os dados, são adotadas algumas métricas, estas detalhadas na seção 4.2.

4.1 DEFINICÃO DE AGRUPAMENTO E CLASSIFICAÇÃO

O agrupamento é um caso especial de classificação. Sua diferença em relação à classificação se dá a capacidade de seus algoritmos em responderem as seguintes questões (Silva, 2003):

- A qual grupo pertence um ponto x do conjunto de dados?
- Quantos grupos existem em um determinado conjunto de dados?
- Quais são esses grupos?

A figura 4.1 sugere uma árvore com os tipos de classificação, onde cada nó da árvore define um modo de resolver o problema de classificação.

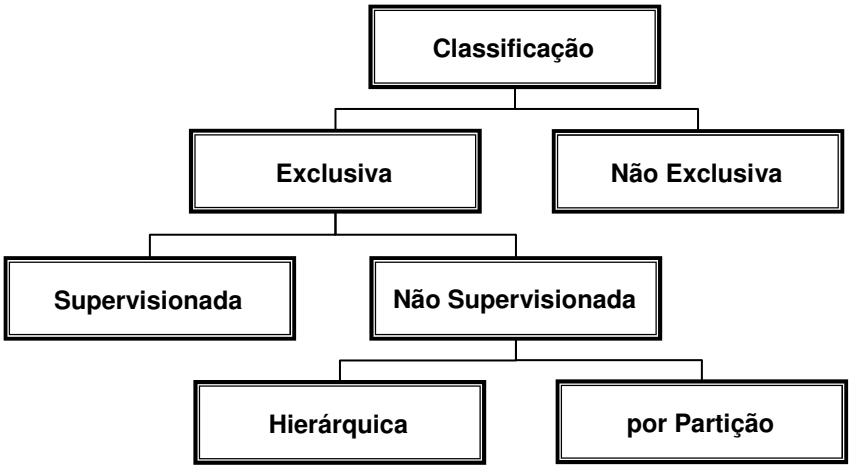

Figura 4.1: Árvore de Tipos de Classificação.

• Classificação exclusiva versus não exclusiva: Uma classificação exclusiva é uma partição de um conjunto X, onde cada ponto pertence exclusivamente a um único grupo. Uma classificação não exclusiva, ou sobreposta, pode assumir que um ponto pertença a vários grupos. Um agrupamento nebuloso é um tipo de classificação não exclusiva, onde o ponto no plano cartesiano é associado a cada um dos grupos com um respectivo grau de pertinência.

Outros termos e notações pertinentes são as seguintes (Silva, 2003):

• Matriz de pertinência: matriz onde o valor de cada elemento é o grau de pertinência de cada ponto aos grupos existentes, sendo representada por uma matriz $U(n \times c)$, onde c é o número de grupos e n representa o número de dados de X. Portanto, cada elemento U_{ki} de $i = 1,...,c$ e $k = 1,...,n$, fornece o grau de pertinência de $x_k \in X$ ao *i*-ésimo grupo.

• Densidade: é a quantidade de dados em relação a alguma unidade do espaço. Para o agrupamento, os grupos são considerados como regiões onde a densidade de dados é maior se comparadas com as regiões vizinhas.

• Centro ou protótipo: é uma referência para o algoritmo determinar se um ponto pertence ou não a um grupo. Um centro denotado por $v_i \in \Re^p$ representa o *i*-ésimo grupo. Cada ponto de X é associado a um grupo de acordo com uma medida de similaridade em relação aos centros dos grupos, de forma que

$$V = \{v_1,...,v_c\} \subset \Re^p \tag{4.1}$$

A equação (4.1) representa o conjunto dos centros dos c grupos.

• Atualização de centros ponto-a-ponto: a atualização dos centros é realizada junto com a avaliação de cada ponto. Neste caso, é possível atualizar o centro mais similar ao ponto analisado, ou todos os centros de acordo com os respectivos graus de pertinência do ponto. Os métodos ponto-a-ponto são necessários quando: (i) os resultados precisam ser conhecidos em tempo real; (ii) os conjuntos de dados são grandes e os métodos por batelada se tornam inviáveis em termos de memória ou tempo de processamento e (iii) os dados são processados de forma contínua o que impossibilita a aplicação do método em batelada.

• Atualização de centros em batelada: os centros são avaliados e/ou atualizados em bloco, ou seja, após a avaliação de todo um conjunto de pontos. Os métodos por batelada são melhores quando se tem interesse apenas no resultado final, isto é, o melhor resultado é obtido no final da execução do algoritmo. Porém, esta técnica torna-se numericamente ou computacionalmente inviável quando o conjunto de dados é grande e a solução do problema é obtida pela minimização de uma função objetivo.

4.2 NORMAS E CRITÉRIOS DE DISTÂNCIA

Com relação as técnicas de agrupamento, um aspecto importante a ser definido é o critério de similaridade ou dissimilaridade entre os pontos, ou seja, as métricas proporcionam meios para essa definição, quantificando o quão próximos ou semelhantes os pontos são. Esta seção tem por objetivo descrever as principais métricas.

4.2.1 Definição de Espaço Métrico

Segundo Domingues (1982), para um conjunto $M \neq 0$ e a função d sendo uma métrica sobre M, as seguintes propriedades são verificadas para quaisquer $x, y, z \in M$, tal que

(i) $d(x, y) = 0 \rightarrow (x = y)$;

(ii) $d(x, y) = d(y, x)$;

(iii) $d(x, y) \leq d(x, z) + d(z, y)$.

Nestas condições, $d(x, y)$ é à distância entre (x, y) e o par (M, d), onde d é uma métrica sobre M, de espaço métrico. Cada objeto de espaço métrico será sempre referido como ponto desse espaço, seja ele um ponto em si mesmo, um número, uma função ou ainda um vetor. A propriedade $d(x, y) \leq d(x, z) + d(z, y)$ é conhecida como desigualdade triangular.

A tabela 4.1 apresenta as mais importantes métricas utilizadas como medidas de distância entre dois pontos.

Tabela 4.1: Métricas Utilizadas como Medidas de Distância

Distância Euclidiana	$d = \sqrt{\sum_{i=1}^{p} (x_i - y_i)^2}$
Distância Euclidiana Média	$d = \sqrt{\frac{\sum_{i=1}^{p} (x_i - y_i)^2}{p}}$
Distância Euclidiana Padronizada	$d = \sqrt{\sum_{i=1}^{p} \left(\frac{x_i - y_i}{s_i} \right)^2}$
Distância Euclidiana Ponderada	$d = \sqrt{(x-y)^T \cdot A(x-y)}$
Distância de Hamming	$d = \sum_{i=1}^{P} w_i \lvert x_i - y_i \rvert$
Distância de Gower	$d = -\log_{10}\left[1 - \frac{1}{p} \sum_{i=1}^{p} \frac{\lvert x_i - y_i \rvert}{x_{\max i} - x_{\min i}} \right]$
Distância de Cattel	$d = \frac{2\left(p - \frac{2}{3}\right) - d_e^2}{2\left(p - \frac{2}{3}\right) + d_e^2}$
Distância Tchebyschev	$d(x,y) = \max\left\{ \lvert x_i - y_i \rvert, ..., \lvert x_p - y_p \rvert \right\}$
Distância de Minkowsky	$d = \left[\sum_{i=1}^{P} w_i \lvert x_i - y_i \rvert^k \right]^{1/k} \rightarrow k \geq 1, x_i \neq y_i$ (a) Para $k = 1$, torna-se a distância Hamming; (b) Para $k = 2,\ w_i = 1,\ i = 1,..., p$, torna-se a distância Euclidiana; (c) Para $k = \infty,\ w_i = 1,\ i = 1,..., p$, torna-se a distância Tchebyschev.

4.2.2 Normas e Normalização

Os algoritmos de agrupamento podem ser sensíveis a escala dos dados e a normalização. Dessa forma, a normalização dos valores dos atributos pode suprir a necessidade de adaptação de escalas com grandezas muito diferentes.

Uma norma sobre um conjunto E é uma função que associa um número real não negativo a cada $u \in E$, indicado por $\|u\|$ e denominado norma de u, de maneira que:

(i) $\|u\| = 0 \rightarrow u = 0$;

(ii) $\|\alpha u\| = |\alpha|\|u\|; \forall \alpha \in \Re; \forall u \in E$;

(iii) $\|u + v\| \leq \|u\| + \|v\|, \forall u, v \in E$.

Se $\|.\|$ é uma norma sobre E, então $d : E \times E \rightarrow \Re$, definida por $d(u,v) = \|u - v\|$ é uma métrica sobre E, pois:

(iv) $d(u,v) = \|u - v\| = 0 \rightarrow u - v = 0 \rightarrow u = v$;

(v) $d(u,v) = \|u - v\| = \|(-1)(v - u)\| = |-1|\|v - u\| = \|v - u\| = d(v,u)$;

(vi) $d(u,v) = \|u - v\| = \|u - w + w - v\| \leq \|u - w\| + \|w - v\| = d(u,w) + d(w,v)$.

A métrica d assim obtida denomina-se métrica induzida pela norma dada sobre E.

A tabela 4.2 apresenta as funções de normalização mais utilizadas. As funções seguintes referem-se a um atributo i para um ponto x_{ki}.

Tabela 4.2: Funções de Normalização.

Função	Valores Normalizados dos Extremos da Escala	Observações
$f(x_{ki}) = \frac{x_{ki} - x_{\min k}}{x_{\max k} - x_{\min k}}$	Máximo = 1 Mínimo = 0	Atributos de Maximização
$f(x_{ki}) = \frac{x_{\max k} - x_{ki}}{x_{\max k} - x_{\min k}}$	Máximo = 1 Mínimo = 0	Atributos de Minimização
$f(x_{ki}) = \frac{x_{ki}}{x_{\max k}}$	Máximo = 1 Mínimo = $\frac{x_{\min k}}{x_{\max k}}$	Atributos de Maximização

Contudo, uma normalização pode ser perigosa para o processo de agrupamento. Por exemplo, se um padrão está presente em um determinado grupo, uma determinada normalização poderá alterar a relação das distâncias entre os pontos e alterar a separação existente entre dois grupos (Silva, 2003), conforme mostrado na figura 4.2.

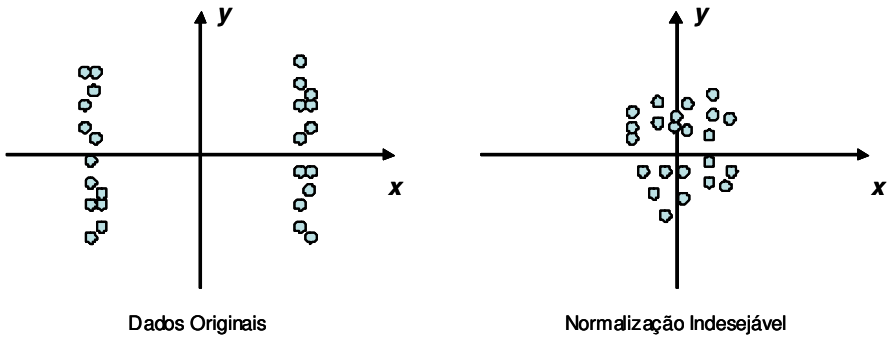

Figura 4.2: Possíveis efeitos de uma normalização.

É possível observar na figura 4.2 que o critério de similaridade e a norma adotada por determinado algoritmo, pode influenciar dramaticamente nos resultados de um agrupamento.

4.3 CATEGORIAS DE ALGORITMOS DE AGRUPAMENTO

Uma multiplicidade de algoritmos de agrupamento é encontrada na literatura. Esses algoritmos podem ser classificados de acordo com várias características. Para Silva (2003), os algoritmos de agrupamento são classificados por características, tais como:

(i) Tipo de Dados**:** Tipo de dado é a capacidade do algoritmo em agrupar ou não, dados de diferente natureza, dados numéricos, caracteres, palavras, imagens, objetos, entre outros.

- Dados Rotulados: são aqueles que assumem valores em um mesmo espaço vetorial multidimensional, acompanhados da classe a que cada dado pertence, podendo haver múltiplas classes, com variâncias e números de dados distintos ou não, para cada classe.
- Dados não Rotulados: são aqueles que assumem valores em um mesmo espaço vetorial multidimensional. O número de classes pode ser conhecido ou não, a variância e o número de dados de cada classe pode diferir ou não. Embora cada um pertença a uma classe específica, não se conhece *a priori* as classes que pertencem.

(ii) Parâmetros do Algoritmo**:** estão relacionados ao conhecimento (*a priori*) das características do conjunto de dados a ser utilizado como número de grupos, tipos de classes, matriz de pertinência, entre outras. Geralmente, estas características ou a falta delas, são fatores importantes que são utilizadas na escolha de um determinado método de agrupamento. Para Gath e Geva (1989), as três maiores dificuldades encontradas durante o agrupamento nebuloso de dados são:

- Número de grupos – nem sempre pode ser definido;
- Características e localização dos centros dos grupos – não são necessariamente conhecidas;
- Grande variabilidade de formas, densidades e número de dados em cada grupo.

Nota-se, portanto, que um dos principais problemas na tarefa de agrupamento é a estimação do número de grupos. Devido a esta característica, a maioria dos algoritmos, sejam eles nebulosos ou não, assume que o número c de classes em um conjunto finito X é conhecido, de modo que, cabe para as técnicas de validação caracterizar o melhor agrupamento.

(iii) Critério de Similaridade ou Dissimilaridade: de acordo com a categoria de critério de similaridade ou dissimilaridade (Jain *et al.*, 1999), os algoritmos podem ser classificados como:

• Agrupamento Hierárquico: Neste tipo, o agrupamento é realizado por sucessivas fusões ou por sucessivas divisões. O método hierárquico é subdividido em métodos aglomerativos e divisivos (Vale, 2005). Os métodos hierárquicos aglomerativos iniciam com os objetos mais similares e são agrupados na formação de um grupo único. Eventualmente o processo é repetido e com o decréscimo da similaridade, todos os subgrupos são fundidos, formando novamente um único grupo com todos os objetos. Os métodos hierárquicos divisivos iniciam com um único subgrupo com todos os objetivos. Este é subdividido em dois subgrupos de tal forma que exista o máximo de semelhança entre os objetos dos mesmos subgrupos e a máxima dissimilaridade entre os elementos de subgrupos distintos. Estes subgrupos são posteriormente subdivididos em outros subgrupos dissimilares. O processo é repetido até que haja tantos subgrupos quantos objetos. Os resultados finais destes agrupamentos podem ser apresentados por gráficos denominados dendrogramas (representados na figura 4.3), que apresentam os dados e os respectivos pontos de fusão ou divisão dos grupos formados em cada estágio. Os dendogramas são impraticáveis para grandes conjuntos de dados (Silva, 2003).

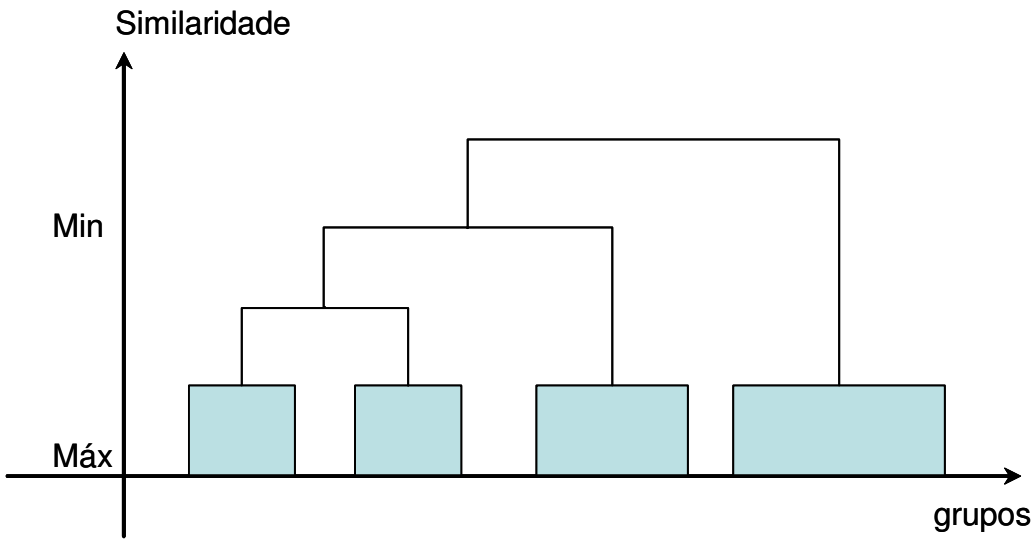

Figura 4.3: Exemplo de um dendograma.

• Agrupamento por partição ou não hierárquico: são agrupamentos onde um número n de dados, determina uma partição em c grupos, ou classes $(c \leq n)$, de modo que os dados que pertencem a um mesmo grupo sejam similares, e os dados de diferentes grupos sejam dissimilares em termos de seus atributos. O valor de c pode ou não ser conhecido *a priori*.

• Agrupamento baseado em densidade: a idéia desta técnica é agrupar pontos vizinhos de um conjunto de dados baseado nas condições de densidade. Os grupos são as regiões densas de pontos, sendo estes grupos separados por regiões de baixa densidade de pontos.

• Agrupamento baseado em grade: sua principal característica é subdividir o espaço em partições menores denominadas de células. Um exemplo de grade bidimensional é ilustrado na figura 4.4. Os algoritmos quantificam o espaço dentro de um número finito de células e realizam todas as operações de quantificação neste espaço.

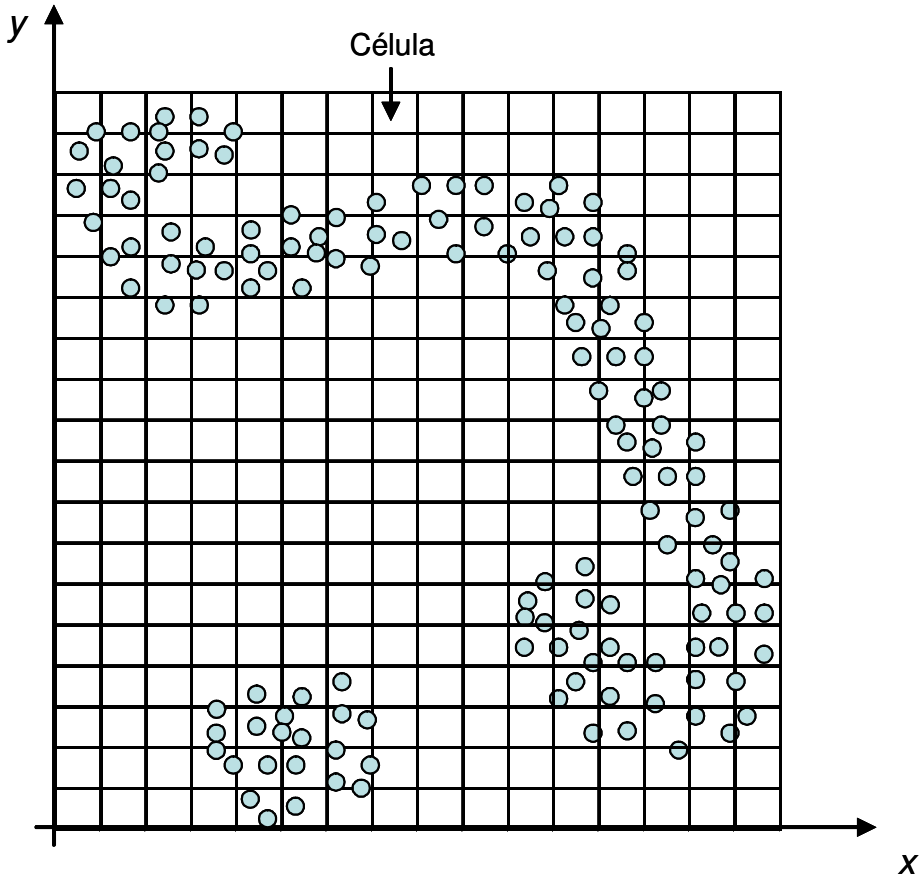

Figura 4.4: Exemplo de grade em um conjunto de dados bidimensional.

(iv) Conceitos e Fundamentos: os algoritmos de agrupamento podem ser caracterizados como:

• Rígidos: são algoritmos que não consideram a sobreposição de grupos aos quais um ponto de conjunto pertence. Do algoritmo resulta um agrupamento com valores na matriz de pertinência restritos ao conjunto $[0, 1]$.

• Estatísticos: são algoritmos onde os conceitos lógico-matemáticos empregados na análise estatística para agrupamento de dados são, principalmente, probabilísticos.

• Nebulosos: são algoritmos que utilizam conjuntos nebulosos para classificar os dados e também consideram que um ponto pode ser classificado em mais de um grupo, mas com diferentes graus de associação.

• Neurais: são algoritmos que adotam abordagens conexionistas (que envolvem quantização vetorial e mapeamento auto-organizável) para o agrupamento. Em geral, utilizam redes neurais artificiais com aprendizagem não supervisionada.

4.4 AGRUPAMENTO NEBULOSO

Os algoritmos de agrupamento nebulosos permitem um grau de associação para cada elemento em cada grupo. Um elemento pertence a diferentes grupos, com diferentes graus de associação, fornecendo informações detalhadas sobre a estrutura de dados. Porém, as maiores dificuldades encontradas durante o agrupamento nebuloso são: o não conhecimento das características e localizações dos centros dos grupos (*a priori*); a grande variabilidade de formas; a densidade e o número de dados em cada grupo; e o fato de o número de grupos nem sempre poderem ser definidos (*a priori*).

Nesta seção será abordado o algoritmo nebuloso – utilizado na identificação dos modelos nebulosos de Takagi-Sugeno – que adota métodos de agrupamento por partição nebulosa, definidos por:

$$U = \begin{bmatrix} w_{11} & w_{12} \dots w_{1c} \\ w_{21} & w_{22} \dots w_{2c} \\ \vdots & \vdots \quad \vdots \\ w_{n1} & w_{n2} \dots w_{nc} \end{bmatrix} \tag{4.11}$$

$$U \in [0,1]^{nc} \left| w_{k,i} \in [0,1], \forall k, i \;\; ; \;\; \sum_{i=1}^{c} w_{k,i} = 1, \forall k \;\; ; \;\; 0 < \sum_{i=1}^{c} w_{k,i} < n, \forall I \right. \tag{4.12}$$

onde U é a matriz com grau de pertinência $w_{k,i}$ e c é o número de grupos.

4.4.1 Algoritmo C-Médias Nebuloso (FCM)

Desenvolvido por Ruspini (1969), o primeiro algoritmo de agrupamento nebuloso é uma extensão do algoritmo C-Médias Rígido (HCM, *Hard C-Means)*, sendo este um dos mais populares métodos de agrupamento de dados (Härdle *et al*., 2003). Ruspini (1969) introduziu a partição nebulosa para descrever estruturas de grupos de um conjunto de dados e sugeriu um algoritmo computacional que determina uma partição nebulosa. Segundo Magalhães (2004), Bezdek generalizou o procedimento de agrupamento de variância mínima para a técnica de agrupamento nebuloso do HCM, criando assim o algoritmo C-Médias nebuloso (FCM, *Fuzzy C-Means*).

O algoritmo FCM procura agrupar os dados criando uma partição em um conjunto de dados, de modo que se minimize a seguinte função objetivo (Scionti e Lanslots, 2005):

$$J_m(X,U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ki}^{m} d_{ki}^{2}, \quad 1 < m < \infty \tag{4.13}$$

onde X é o conjunto de dados, U é a matriz de pertinência e c é o número de grupos; $d_{ki} = \|x_k - v_i\|$ onde v_i é o centro da *i*-ésima classe de grupos; w_{ki} denota o grau de pertinência de x_k na classe i; e m é um valor que modula o quão nebulosa é a partição obtida.

Este algoritmo adota a distância Euclidiana como medida de similaridade entre o dado e o centro do grupo. O desempenho do FCM é favorecido quando os conjuntos de dados são separáveis ou quando os grupos têm aproximadamente os mesmos tamanhos e formas, porém ele não correlaciona os valores dos atributos entre os dados.

Etapas do algoritmo FCM:

1. Escolher o número de grupos $c, 2 \le c \le n$, o parâmetro $m > 1$, o critério de parada $\varepsilon > 0$, o número máximo $l_{\max}$ de iterações. Iniciar U^0 aleatoriamente e o contador $l = 1$;
2. Calcular os c centros das classes $v_i^l \quad i = 1,...,c$ com U^l e a fórmula:

$$v_i^l = \frac{\sum_{k=1}^{n} [u_{ki}^{l-1}]^m x_k}{\sum_{k=1}^{n} [u_{ki}^{l-1}]^m}, i = 1,...,c. \qquad (4.14)$$

3. Atualizar a matriz de pertinência:
   Para $1 \le i \le c$, $1 \le k \le n$;
   Se $d_{ki} > 0$ então

$$u_{ki}^l = \left[\sum_{j=1}^{c} \left(\frac{d_{ki}}{d_{kj}} \right)^{\frac{1}{m-1}} \right]^{-1} \qquad (4.15)$$

   Senão
   Se $d_{ki} = 0$ para $i \in I \le c$, então
   Definir u_{ki}^l para $i \in I$ com um número real positivo que satisfaça a condição: $\sum u_{ki}^l = 1$ deste modo:
   $u_{ki}^l = 1 - \sum u_{ki}^l$
   Definir $u_{ki}^l = 0$ para $i \in c - I$;
4. Calcular $\Delta = \left\| U^l - U^{l-1} \right\| = \max_{ji} \left| u_{ji}^l - u_{ji}^{l-1} \right|$, $j = 1,...,n$, $i = 1,...,c$;
5. Se $\Delta > \varepsilon$ e $l < l_{\max}$, $l = l + 1$ e retornar para o passo (2); Senão parar.

Algoritmo 4.1: C-Médias Nebuloso (FCM).

4.4.2 Algoritmo Gustafson-Kessel (GK)

O algoritmo proposto por Gustafson-Kessel é uma extensão do algoritmo FCM, empregando uma norma adaptável da distância, a fim de detectar grupos de formas geométricas diferentes na série de dados (Guerra, 2006). Utiliza-se no algoritmo GK a distância Euclidiana ponderada – que remove algumas limitações da distância Euclidiana – por não depender das escalas dos atributos e correlacionar atributos com escalas diferentes (Krishnapuram e Kim, 1999), dado por

$$d_{ki}{}^{2} = (x_k - v_i)^T M_i (x_k - v_i), \qquad 1 \le i \le; \qquad 1 \le k \le n \tag{4.16}$$

onde M_i é a matriz de covariância para o grupo i, simétrica e positiva, sendo determinada por:

$$M_i = \left[\left(\det(F_i)^{\frac{1}{n+1}} \right) F_i^{-1} \right] \tag{4.17}$$

$$F_i = \frac{\sum_{k=1}^{\mathrm{n}} \mu_{ki}^{(l-1)} (x_k - v_i^l)(x_k - v_i^l)^{\mathrm{T}}}{\sum_{k=1}^{\mathrm{n}} \mu_{ki}^{(l-1)}} \tag{4.18}$$

$$F_i = \frac{\sum_{k=1}^{\mathrm{n}} [\mu_{ki}^{l-1}]^{\mathrm{m}} (x_k - v_i^l)(x_k - v_i^l)^{\mathrm{T}}}{\sum_{k=1}^{\mathrm{n}} [\mu_{ki}^{l-1}]^{\mathrm{m}}} \tag{4.19}$$

onde μ_{ik} é o grau de pertinência do objeto i do conjunto de dados em relação ao agrupamento k e m é o coeficiente nebuloso.

A substituição das equações (4.17) e (4.19) na equação (4.16) resulta em uma norma quadrada generalizada da distância de Mahalanobis entre o x_k e o seu grupo v_i, onde a covariância seja relacionada pelos graus em U (Guerra, 2006).

O procedimento desse agrupamento é um processo interativo, na qual os centros dos grupos são movidos no espaço (entrada-saída) até que algum critério de convergência seja encontrado (Van Lith *et al.*, 2002).

Etapas do algoritmo GK:

1. Escolher o número de grupos $c, 2 \le c \le n$, o parâmetro $m > 1$, o critério de parada $\varepsilon > 0$, o número máximo $l_{\max}$ de iterações. Iniciar U^0 aleatoriamente e o contador $l = 1$.
2. Calcular os c centros das classes $v_i^l \quad i = 1,...,c$ com U^l e a fórmula (4.12):
3. Calcular a matriz de covariância usando (4.19) para $\forall i$, $i = 1,2,...,c$;
4. Calcular a distância de Mahalanobis:

$$d_{ki} = (x_k - v_i^l)^T \left[\left(\det(F_i)^{\frac{1}{n+1}} \right) F_i^{-1} \right] (x_k - v_i^l), \qquad (4.20)$$

$i = 1,2,...,c, \quad k = 1,2,...,n$

5. Atualizar a matriz de pertinência:
   Para $1 \le i \le c$, $1 \le k \le n$,
   Se $d_{ki} > 0$ então
   Atualizar u_{ki}^l usando a fórmula (4.15)
   Senão
   Se $d_{ki} = 0$ para $i \in I \le c$, então
   Definir u_{ki}^l para $i \in I$ com um número real positivo que satisfaça a condição: $\sum u_{ki}^l = 1$ deste modo: $u_{ki}^l = 1 - \sum u_{ki}^l$
   Definir $u_{ki}^l = 0$ para $i \in c - I$.
6. Calcular $\Delta = \left\| U^l - U^{l-1} \right\| = \max_{ji} \left| u_{ji}^l - u_{ji}^{l-1} \right|$, $j = 1,...,n$. $i = 1,...,c$,
7. Se $\Delta > \varepsilon$, $l < l_{\max}$ e $l = l + 1$, retornar ao passo (2); Senão parar.

Algoritmo 4.2: Gustafson-Kessel (GK).

4.4.3 Algoritmo K-Médias Nebuloso (FKM)

O algoritmo K–Médias nebuloso (FKM, *Fuzzy K-Means*) é um método de classificação que se apresenta como uma extensão do algoritmo k–Médias e envolve um processamento simples de estimação de parâmetros (Guerra, 2006).

O algoritmo trabalha a partir de um procedimento iterativo onde os objetos, inicialmente em posição aleatória, são classificados em k classes (ou grupos). Dado o número de classes desejadas, são calculados os centros de cada classe com base na média dos atributos dos objetos. Num próximo passo, os objetos são realocados entre as classes, de acordo com a similaridade entre eles. A medida de similaridade pode ser uma métrica de distância, como a Euclidiana ou diagonal.

Seja o conjunto de pontos $X = \{x_1, x_2, \ldots, x_n\}$ e o conjunto de funções de pertinência $U_{ki}(x) \rightarrow [0,1],\ i = 1, \ldots k$ que definem os graus de pertinência de cada x_k com relação a cada uma das k classes. Os centros de cada uma das classes podem ser calculados pela equação (4.14), onde $m,\ 1 < m \leq \infty$, é um coeficiente que vai ponderar o quanto o grau de pertinência influencia na métrica de distância definida.

A atualização dos graus de pertinência a cada iteração é dada pela equação (4.15), onde d é a métrica de distância empregada.

O algoritmo finaliza quando um determinado número de iterações é alcançado ou quando a matriz $U = < A_{ki} >$ é menor que certo limiar δ de convergência.

E possível observar que no K-Médias têm-se cada elemento pertencendo a apenas um grupo, resultando, portanto, em grupos disjuntos. Já no K-Médias nebuloso cada elemento apresenta um grau de pertinência, segundo uma função, com relação a cada grupo. Desta forma, a saída do algoritmo é um agrupamento nebuloso, onde cada grupo é um conjunto nebuloso de todos os elementos.

Esta técnica foi aplicada em várias áreas, incluindo a compressão de imagem e de dados, processamento de dados para modelamento de sistemas que usam redes neurais funções de base e decomposição da tarefa nas arquiteturas de redes neurais heterogêneas (Hammouda e Karray, 2000).

Etapas do algoritmo FKM:

1. Fase de Iniciação;
   Escolher o número de grupos $c, 2 \le c \le n$, o parâmetro $m > 1$, o critério de parada $\varepsilon > 0$, o número máximo $l_{\max}$ de iterações. Iniciar U^0 aleatoriamente e o contador $l = 1$. Escolher os valores padrões para η e a.
   Determinar o centro inicial da classe, v_1^0;
2. Para cada iteração $l = 1,2,...$;
   2.1 Para $l \le i \le c$, $l \le k \le n$;
   2.1.1 Calcular U^l utilizando (4.15);
   2.2 Atualizar v_i^l, utilizando (4.14);
3. Calcular $\Delta = \left\| U^l - U^{l-1} \right\| = \max_{ji} \left| u_{ji}^l - u_{ji}^{l-1} \right|$, $j = 1,...,n$. $i = 1,...,c$;
4. Se $\Delta > \varepsilon$ e $l < l_{\max}$, $l = l + 1$ e retornar para o passo (2); Senão parar.

Algoritmo 4.3: K-Médias Nebuloso (FKM).

4.4.4 Algoritmo Gath-Geva Modificado (MGG)

Uma desvantagem na construção de modelos nebulosos do tipo Takagi-Sugeno é a matriz de covariância (4.17) poder ter elementos não nulos fora da diagonal e conseqüentemente ocorrer erro de decomposição do conjunto nebuloso. Para preservar o particionamento no espaço do antecedente, transformações lineares nas variáveis de entrada podem ser utilizadas, complicando a interpretação das regras. Para contornar este problema, o algoritmo Gath-Geva é modificado, baseando-se na maximização da esperança de modelos Gaussianos, sendo este muito utilizado para estimar a densidade de probabilidade de um conjunto de dados (Abonyi *et al.*, 2005). Para o algoritmo GG modificado (MGG, *Modified Gath-Geva*), usa-se a mesma função objetivo (4.13), onde, cada grupo de dados contém uma distribuição na entrada, um modelo local e uma distribuição na saída:

$$p(\phi, y) = \sum_{i=1}^{c} p(\phi, y, n_i) = \sum_{i=1}^{c} p(y|\, \phi, n_i)\; p(x|\, n_i)\; p(n_i) \qquad (4.21)$$

onde $p(n_i)$ é a possibilidade *a priori* dos dados pertencentes ao grupo i; $p(x|\, n_i)$ é a distribuição na entrada; e $p(y|\, \phi, n_i)$ é a distribuição na saída.

Etapas do algoritmo MGG:

1. Fase de Iniciar parâmetros;
   Escolher o número de grupos $c, 2 \le c \le n$, o parâmetro $m > 1$ e o critério de parada $\varepsilon > 0$. Iniciar usando (4.12).
2. Repetir para cada iteração $l = 1,2,...$;
3. Calcular o centro dos grupos usando (4.14);
4. Computar a medida de distância por d_{ki}^2. A distância do protótipo é baseada na matriz de covariância dada por (4.18) para $1 \le i \le c$.
5. A função da distância é escolhida por:

$$d_{ki}^2(x_k, v_i) = \frac{(2\pi)\left(\frac{n+1}{2}\right)\sqrt{\det(F_i)}}{\frac{1}{n}\sum_{k=1}^{n}\mu_{ik}} \exp\left(\frac{1}{2}(x_k - v_i^l)^T F_i^{-1}(x_k - v_i^l)\right) \qquad (4.22)$$

6. Atualizar a matriz usando a equação (4.15) e calcular $1 \le i \le c$ e $1 \le k \le n$ até $\Delta = \left\| U^l - U^{l-1} \right\| < \varepsilon$.

Algoritmo 4.4: Gath-Geva Modificado (MGG).

4.4.5 Algoritmo do Modelo de Misturas Gaussianas (GMM) usando a Maximização da Esperança (EM)

A densidade de mistura é definida como a média ponderada da densidade de componentes (Verbeek, 2004). A densidade de mistura Gaussiana completa, denominada λ, é parametrizada por

$$\lambda = \{w_m, \mu_m, \Sigma_m\}, m = 1, \ldots Z \tag{4.23}$$

onde μ_m é o vetor de médias, Σ_m é a matriz de covariância, W_m é o coeficiente de ponderação de cada componente e Z é o número de componentes gaussianas.

Dessa forma, a mistura de densidade de probabilidades Gaussianas é uma soma ponderada de M densidades, definida pela equação,

$$p(x|\lambda) = \sum_{m=1}^{Z} w_m b_m(x) \tag{4.24}$$

onde x é um vetor aleatório de ordem D e $b_m(x)$ são as densidades componentes.

Cada densidade componente é uma função Gaussiana de dimensão D da forma, tal que

$$b_m(x) = \frac{1}{(2\pi)^{D/2} |\Sigma_m|^{1/2}} e^{\left\{-\frac{1}{2}(x-\mu_m)^T \Sigma_m^{-1}(x-\mu_m)\right\}} \tag{4.25}$$

onde $|.|$ indica determinante e $(\ )^T$ representa a transposição. Neste contexto, a ponderação das misturas satisfaz a condição $\sum_{m=1}^{Z} w_m = 1$.

O GMM pode apresentar três tipos de matriz de covariância: (i) distribuída a cada componente Gaussiana; (ii) uma única matriz para todas as componentes Gaussianas de um dado modelo e (iii) uma única matriz de covariância para todas as componentes de todos os modelos.

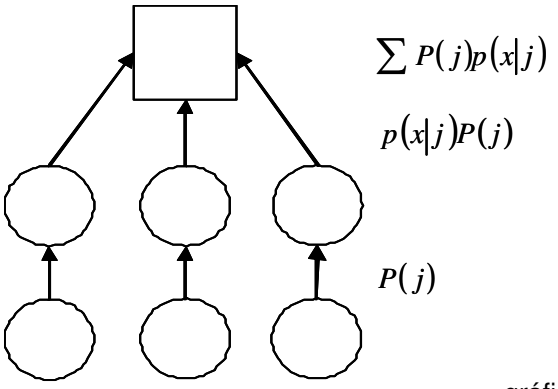

Figura 4.5: Representação gráfica para um GMM.

Os componentes da mistura têm a seguinte forma genérica:

$$p(x|j) = \frac{1}{(2\pi)^{D/2}\|C_j\|^{1/2}} e^{\left\{-\frac{1}{2}(x-\mu_j).C_j^{-1}(x-\mu_j)\right\}} \tag{4.26}$$

As matrizes de covariância C_j podem apresentar diversas formas que influenciam no número de componentes necessários para a representação de uma densidade:

- Esférica: Nesta escolha as covariâncias são descritas por apenas um parâmetro $C_j = \sigma_j^2 I$;
- Elíptica: Cada matriz de covariância é descrita por N parâmetros, onde N é a dimensão do espaço de dados, assim $C_{kl,j} = \delta_{kl}\sigma_{k,j}$.
- Completa: A covariância utilizada tem $\frac{N^2+N}{2}$ parâmetros, ou seja, $C_{kl,j} = \sigma_{kl,j}^2$.

Existem vários métodos de estimação dos parâmetros do GMM. Um método que apresenta bons resultados é a estimação da máxima verossimilhança - ML (Cirigliano, 2007). Para um conjunto de dados de treinamento, a estimação ML tenta encontrar os parâmetros do modelo que maximizem a verossimilhança do GMM. Para uma

seqüência de vetores de características $X = \{x_1, \ldots x_T\}$, supondo independência entre os mesmos, a verossimilhança do GMM é dada por,

$$p(x|\lambda) = \prod_{t=1}^{T} p(x_t|\lambda) \tag{4.27}$$

Normalizando pelo número total de vetores T e usando o logaritmo, tem-se,

$$\log p(x|\lambda) = \frac{1}{T} \prod_{t=1}^{T} \log p(x_t|\lambda) \tag{4.28}$$

onde x_t é o *t*-ésimo vetor de X.

Esta expressão é uma função não-linear de parâmetros λ e cuja maximização direta não é de fácil implementação. Entretanto, a estimação dos parâmetros obtidos pelo método ML pode ser conseguida iterativamente, utilizando-se um caso especial do algoritmo da maximização da esperança - EM (Cirigliano, 2007).

Em geral, os parâmetros descrevem as características de uma população. Seus valores são estimados de amostras coletadas dessa população. O algoritmo de aprendizado por maximização da esperança (EM) é uma técnica geral para estimação de máxima verossimilhança dos parâmtros, ou seja, estima os parâmetros que sejam os mais consistentes como os dados da amostra no sentido de maximizar a função de verossimilhança (Luna, 2004), aplicado principalmente a aprendizado não-supervisionado, isto é, agrupamento e estimação da mistura de densidades (Lima, 2004).

O algoritmo EM tem por meta aumentar a estimação do logaritmo da verossimilhança de dados completos. Em síntese, cada iteração do algoritmo EM é definido por dois passos combinados em um único algoritmo (Moon, 1996), decritos por:

(i.) Passo E – *Expectation*: são encontrados os valores esperados das estatísticas suficientes para os dados completos, dado os dados incompletos e as estimativas

atuais dos parâmetros. Usando os dados observáveis e o modelo atual, o passo E calcula o valor esperado do logaritmo da verossimilhança de dados completos a cada iteração, na forma:

$$Q\left(\theta,\theta^{(k)}\right)=E\left[l_c(\theta;T)|\chi\right] \tag{4.29}$$

onde o sub-índice c implica que a verossimilhança está sendo avaliada junto aos dados completos. O símbolo sobrescrito k refere-se aos parâmetros da *k*-ésima iteração do algoritmo. O passo E produz uma função determinística Q dos parâmetros do modelo.

(ii.) Passo M – *Maximization*: são utilizadas estas estatísticas suficientes para fazes uma estimativa de máxima verossimilhança. Dessa forma, o passo M maximiza a função Q com respeito aos parâmetros do modelo produzindo:

$$Q^{(k+1)}=\arg\max_{\theta} Q\left(\theta,\theta^{(k)}\right) \tag{4.30}$$

O algoritmo EM realiza uma estimativa dos parâmetros no sentido de aumentar o valor da função Q a cada iteração. A função Q, portanto, vai corresponder ao valor esperado do logaritmo da verossimilhança dos dados completos.

A cada iteração do algoritmo EM, as seguintes fórmulas de reestimação são usadas para a modelagem da *m*-ésima Gaussiana, as quais garantem um crescimento monotônico do modelo de verossimilhança:

$$\overline{w}_m=\frac{1}{T}\sum_{t=1}^{T} p\left(m|x_t,\lambda\right) \tag{4.31}$$

$$\overline{\mu}_m=\frac{\sum_{t=1}^{T} p\left(m|x_t,\lambda\right)x_t}{\sum_{t=1}^{T} p\left(m|x_t,\lambda\right)} \tag{4.32}$$

$$\overline{\sigma}_{mj}^2 = \frac{\sum_{t=1}^{T} p(m|x_t,\lambda) x_{tj}^2}{\sum_{t=1}^{T} p(m|x_t,\lambda)} - \mu_{mj}^2 \tag{4.33}$$

onde $\overline{\sigma}_{mj}^2$, x_{tj} e $\overline{\mu}_{mj}$, $j = 0,\ldots,D$, $m = 1,\ldots,M$ referem-se aos elementos dos vetores $\overline{\sigma}_m^2$, x_t e $\overline{\mu}_m$, respectivamente, e $p(m|x_t,\lambda)$ é a probabilidade a posteriori para uma classe m dada por,

$$p(m|x_t,\lambda) = \frac{w_m b_m(x_t)}{\sum_{k=1}^{M} p_k b_k(x_t)} \tag{4.34}$$

Os passos E e M são repetidos até que o processo de maximização não produza nenhum melhoramento significativo, ou seja, quando a função Q é ontínua, o algoritmo EM converge para um ponto estacionário da função de verossimilhança. Quando esta função tem um único máximo, EM converge para esta estimativa de máxima verossimilhança global (Luna, 2004). As equações obtidas são utilizadas para calcular a covariância, a média e os parâmetros do modelo (Ye e Spetsakis, 2003).

O agrupamento K-médias é um exemplo clássico de aplicação do algoritmo EM. Ele pode ser visto como um problema de estimação das médias do modelo de misturas Gaussianas K. Um pressuposto do agrupamento K-médias é que suas probabilidades de mistura são iguais e cada distribuição Gaussiana tem a mesma variância (Ye e Spetsakis, 2003).

Etapas do algoritmo GMM – EM:

1. Fase de Iniciação;
   Os N dados são divididos em k grupos com o mesmo número de componentes;
2. Calcular as médias (centros) para cada grupo: $\mu_j = \frac{1}{N_j}\sum_{n\in S_j} x_n$ (4.35);
3. Calcular as distâncias de cada dado em relação a cada um dos centros e reagrupar os dados no grupo de centro mais próximo de forma a minimizar $E = \sum_{j=1}^{K}\sum_{n\in S_j}\left\|x_n - \mu_j\right\|^2$ (4.36);
4. Atualizar a matriz usando as equações (4.29), (4.30) e (4.34);
5. Se Q não é contínuo, retornar para o passo (2). Senão calcular a covariância e parar.

Algoritmo 4.5: GMM - EM.

4.5 ESTIMAÇÃO DOS PARÂMETROS DO CONSEQÜENTE DO MODELO TS

A estimação é a determinação de grandezas físicas não observáveis a partir de grandezas mensuráveis. A teoria de estimação compreende basicamente duas classes de problemas:

(i) Identificação Experimental: é a determinação dos parâmetros de um modelo de um sistema através de medidas das suas entradas e saídas. Esta fora de identificação é conhecida também como “modelagem caixa preta";
(ii) Estimação de Estados: é a determinação dos estados de um sistema a partir de medidas das suas entradas e saídas.

Para motivar a discussão do problema de estimação de parâmetros, supomos que estejam disponíveis medições de duas grandezas x e y. Tais variáveis estão

relacionadas da seguinte forma: $y = f(x)$. Uma situação comum ocorre quando a função $f(.)$ é caracterizada por um vetor de parâmetros ϕ. Nesse caso, diz-se que $f(.)$ é parametrizada por θ e tal relação pode ser explicitamente representada escrevendo-se $y = f(x, \theta)$. O problema de estimação dos parâmetros consiste em estimar θ a partir de um conjunto de medidas de x e y (Aguirre, 2000).

Importantes problemas na área de identificação ocorrem porque as medidas de entrada e saída podem estar corrompidas por ruídos. Para baixos níveis de ruído, o método de mínimos quadrados pode produzir uma excelente estimação dos parâmetros do sistema. Porém em certas situações, o ruído ou erro na equação de regressão não é branco, se apresentando como um ruído autocorrelacionado (colorido).

De forma a contornar este problema (apesar de serem inevitavelmente dependentes da precisão do modelo do ruído) podem ser utilizados os estimadores estendidos de mínimos quadrados, os estimadores de mínimos quadrados generalizados e os métodos de erro de predição (Almeida, 2005).

4.5.1 Método de Mínimos Quadrados

Os estudos de astronomia que tiveram origem na Babilônia em 300a.C. estimularam a invenção e o desenvolvimento do método dos mínimos quadrados, publicado em 1806 por Legendre. Mas o conceito básico da técnica dos mínimos quadrados foi elaborado por Carl Friedrich Gauss, matemático e astrônomo alemão que enunciou $y = a + b(x)$ (Pucciarelli, 2005). Uma forma de definir um conjunto de equações é considerar uma função escalar $y = f(x)$ e várias aplicações dessa mesma função, tal que

$$\begin{aligned} y_1 &= f(\mathrm{x}_1) \\ y_2 &= f(\mathrm{x}_2) \\ &\vdots \\ y_N &= f(\mathrm{x}_\mathrm{N}) \end{aligned} \tag{4.37}$$

No caso vetorial, $f(\mathrm{x}):\Re^{\mathrm{n}} \to \Re$ depende de um vetor de n parâmetros e θ sendo parametrizada por $\theta \in \Re^{n}$, sendo representados da seguinte forma:

$$y = f(\mathrm{x},\theta) \tag{4.38}$$

A função (4.38) define um conjunto de equações, bastando para isso escrever (4.38) para várias observações do escalar y (variável dependente) e do vetor de variáveis independentes da seguinte forma:

$$\begin{aligned} y_1 &= f(\mathrm{x}_1,\theta) \\ y_2 &= f(\mathrm{x}_2,\theta) \\ &\vdots \\ y_N &= f(\mathrm{x}_N,\theta) \end{aligned} \tag{4.39}$$

A função (4.38) define uma família de equações, sendo que N membros de tal família estão representados em (4.39). Sendo y_i a ***i***-ésima observação de y e $x_i = [x_{1i}, x_{2i}, \ldots, x_{ni}]^T$ são as ***i***-ésimas observações dos n elementos do vetor x, a equação (4.36) pode ser reescrita como:

$$y = x^T\theta \tag{4.40}$$

onde implica f ser linear nos parâmetros. Se f não satisfizer a tal condição, θ poderá em princípio ser estimada por métodos de estimação.

Podemos supor que se conhece o valor estimado do vetor de parâmetros $\hat{\theta}$, e que é cometido um erro ε ao se tentar explicar o valor observado y a partir do vetor de regressores x e de $\hat{\theta}$, ou seja:

$$y = x^T\hat{\theta}\varepsilon \tag{4.41}$$

Em relação ao erro ε , a seguinte distinção será realizada:

• Ruído: corrompe a variável a ser explicada y , devido a certos fatores como erros de medição, não linearidades, dinâmica não modelada, entre outros. Será representado pela variável e da seguinte forma:

$$y = x^T \hat{\theta} e \tag{4.42}$$

Porém, vale salientar que alguns autores denominam de e o erro de regressão, reservando o termo ruído exclusivamente para ruído de medição.

• Resíduo: caracterizado como o erro cometido pelo modelo ao explicar a variável y com os regressores x e os parâmetros estimados $\hat{\theta}$. Será representado pela variável ξ da seguinte forma:

$$y = x^T \hat{\theta} \xi \tag{4.43}$$

Representando a equação (4.43) em forma matricial, tem-se:

$$y = X \hat{\theta} \xi \tag{4.44}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N\text{x}1} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}_{N\text{x}n} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{bmatrix}_{n\text{x}1} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{bmatrix}_{N\text{x}1}$$

onde $x_i = [x_{1i}, x_{2i}, \ldots, x_{ni}]^T$ são as ***i***-ésimas observações $(i = 1, \ldots, N)$ dos n elementos do vetor x. Considerando a existência do erro (resíduo), $\xi = y - X\hat{\theta}$, o problema de mínimos quadrados é determinar $\hat{\theta}$ que minimize o critério J_{MQ}, que é um índice que quantifica a qualidade do ajuste de $X^T \hat{\theta}$ ao vetor de dados y , onde

$$J_{\mathrm{MQ}} = \sum_{i=1}^{\mathrm{N}} \xi(i)^2 = \xi^T \xi = \left\| \xi \right\|^2 \tag{4.45}$$

substituindo $\xi = y - X\hat{\theta}$ em (4.45), tem-se

$$J_{\mathrm{MQ}} = (y - X\hat{\theta})^T (y - X\hat{\theta}) = y^T y - y^T X\hat{\theta} - \hat{\theta}^T X^T y + \hat{\theta}^T X^T X\hat{\theta} \tag{4.46}$$

Para minimização da forma quadrática, é necessário resolver: $(\partial JMQ / \partial \hat{\theta}) = 0$.

$$(\partial J_{\mathrm{MQ}} / \partial \hat{\theta}) = -(y^T X)^T - X^T y + 2X^T X\hat{\theta} \tag{4.47}$$

Deste modo pode-se escrever a equação normal:

$$X^T X\hat{\theta} = X^T y \tag{4.48}$$

$$\hat{\theta} = [X^T X]^{-1} X^T y \tag{4.49}$$

onde $[X^T X]^{-1} X^T$ é conhecida como matriz pseudo-inversa de X. A equação (4.49) é o estimador que fornece o valor de $\hat{\theta}$ que minimiza o somatório do erro quadrático,

$$\theta_{\mathrm{MQ}} = \arg\phi \min J_{\mathrm{MQ}} = [X^T X]^{-1} X^T y \tag{4.50}$$

O $\arg\phi \min J_{\mathrm{MQ}}$, indica o argumento pertencente ao domínio de θ, que minimiza a função objetivo, J_{MQ}. A equação (4.49) é conhecida como o estimador de mínimos quadrados clássico ou ordinário.

4.6 MÉTODOS DE VALIDAÇÃO DE GRUPOS

A validação é o procedimento final da identificação de sistemas, sendo efetuada após a determinação da estrutura e estimação de parâmetros, e tem como objetivo verificar se o modelo matemático obtido é capaz de representar a dinâmica do sistema em questão, e também se o modelo é não polarizado. Um modelo polarizado pode ser capaz de prever a resposta do sistema quando utilizado o mesmo conjunto de dados da estimação, mas isto pode não ocorrer para dados desconhecidos (Guerra, 2006).

Quando um algoritmo não considera um critério de validação, utiliza-se de métodos para determinar o número de grupos e a partição mais adequada dentre os resultados encontrados. Estes métodos utilizam funções de validação aplicadas sobre a partição. As funções de validação têm sido usadas como indicadores da qualidade do resultado da partição, pois fornecem índices de validação. Estes índices podem ser considerados como um valor que disponibiliza uma maneira de avaliar os resultados do agrupamento.

Os critérios solicitados pelo índice de validação para definir uma partição aceitável baseiam-se nos três requisitos citados a seguir (Gath e Geva, 1989): (i) separação entre grupos resultantes; (ii) determinada concentração de pontos em torno do centro de um grupo e (iii) menor número de grupos possível, desde que obedeça aos requisitos anteriores. São apresentados na tabela (4.3) alguns índices de validação para partições nebulosas (Silva, 2003).

Tabela 4.3: Funções de Validação.

Índice	**Função de Validação**	**F(.)**
Partitioning Entropy (PE)	$PE(U,c) = -\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki}\log_a u_{ki}$, $1 < a < \infty$	$F(\downarrow)$
Partitioning Coefficient (PC)	$PC(U,c) = \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c} {u_{ki}}^2$	$F(\uparrow)$
Hipervolume Nebuloso (HPV)	$HPV = \sum_{i=1}^{c}\left[\det(F_i)\right]^{1/2}$ F_i (equação 4.19), $\forall_i$, $i = 1,2,...,c$	$F(\downarrow)$
Partitioning Average Density (AD)	$AD(k) = \frac{1}{k}\sum_{i=1}^{k}\frac{\sum_{k=1}^{n} u_{ki}}{\left[\det(F_i)\right]^{1/2}}$ $\forall x_k \in \{x_k \mid (x_k - v_i)^T F_i^{-1}(x_k - v_i) < 1\}$	$F(\uparrow)$
Partitioning Density (PD)	$PD = \frac{\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki}}{HPV}$ $\forall x_k \in \{x_k \mid (x_k - v_i)^T F_i^{-1}(x_k - v_i) < 1\}$	$F(\uparrow)$
Xie-Beni (XB)	$XB = \frac{\sum_{k=1}^{c}\sum_{i=1}^{n} \boldsymbol{u}_{ki}^{m}\lVert v_i - x_k\rVert^2}{n(\min_{i\neq k}\lVert v_i - x_k\rVert^2)}$	$F(\downarrow)$
Fukuyama-Sugeno (FS)	$FS_m = \sum_{k=1}^{c}\sum_{i=1}^{n} \boldsymbol{u}_{ki}^{m}\left(\lVert x_k - v_i\rVert_A^2 - \lVert v_i - v\rVert_A^2\right)$	$F(\downarrow)$
Legenda		
$F(\downarrow)$: Minimiza a função	$F(\uparrow)$: Maximiza a função	

O Erro Médio Quadrático (MSE, *Mean Square Error*) é um índice muito utilizado na validação de modelos. O problema é que este índice não proporciona uma resposta explícita em relação a real qualidade do modelo obtido. Então para isso, utiliza-se nesse trabalho outro índice para avaliar o desempenho do modelo: a média harmônica $\left(R^2{}_{mh}\right)$ do coeficiente de correlação múltipla $\left(R^2\right)$ das etapas de estimação $\left(R^2{}_{est}\right)$ e validação $\left(R^2{}_{val}\right)$.

O coeficiente de correlação múltipla é calculado por

$$R^2 = 1 - \frac{\sum_{t=1}^{K}\left(y(t) - \hat{\bar{y}}(t)\right)^2}{\sum_{t=1}^{K}\left(y(t) - \bar{y}\right)^2} \tag{4.51}$$

onde K é o número de amostras avaliado, $y(t)$ é a saída real do processo, $\hat{y}(t)$ e a saída estimada pelo modelo de TS, $\bar{y}$ é a média das medidas do sistema e R^2 indica a aproximação do modelo aos dados medidos do processo.

A média harmônica dos números reais positivos $a_1,\ldots,a_n$ é definida como sendo o número de membros n dividido pela soma do inverso dos membros, como segue

$$R^2{}_{mh} = \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{an}} \tag{4.52}$$

A média harmônica nunca é maior do que a média geométrica ou do que a média aritmética. Outra forma de calcular a média harmônica de dois números é multiplicar os dois números e dividir o resultado pela média aritmética dos dois números. Matematicamente:

$$R^2{}_{mh} = \frac{\alpha \cdot \beta}{\left(\frac{\alpha + \beta}{2}\right)} \tag{4.53}$$

Essa fórmula é equivalente à primeira, mas mais simples em alguns casos. Quando o valor de R^2 é igual a 1, indica uma exata adequação do modelo para os dados medidos do processo. Para aplicações práticas, o valor de R^2 é considerado suficiente no intervalo $0{,}90 \leq R^2 \leq 1{,}00$ (Coelho e Alves, 2006).

A complexidade computacional do modelo está relacionada ao número de regras utilizadas na determinação do resultado do modelo nebuloso de TS.

4.7 RESUMO DO CAPÍTULO

Foram abordados alguns dos principais conceitos sobre agrupamento nebuloso de dados e os algoritmos de agrupamento que podem ser utilizados para a determinação das funções de pertinência – conjuntos nebulosos – na parte do antecedente e a utilização do método de mínimos quadrados para determinação das expressões funcionais da parte do conseqüente do modelo nebuloso TS. No entanto, deve-se enfatizar que apesar dos avanços apresentados na recente literatura, nenhuma abordagem é aceita como unanimidade.

No próximo capítulo são apresentados os estudos de casos de séries temporais a serem analisados pelos algoritmos de agrupamento de dados nebulosos combinados ao modelo TS na tarefa de previsão.

Capítulo 5

ESTUDO DE CASOS

Com a finalidade de analisar o comportamento do setor de agronegócios brasileiro, utilizam-se duas séries temporais econômicas, relativas ao preço do álcool hidratado e o grão de soja. A escolha dessas *commodities* ocorreu devido a importância de ambos para o PIB brasiliero (PIB, Produto Interno Bruto).

O surgimento de novas tecnologias de motores e veículos (flex-fuel) tem provocado mudanças importantes na tradicional postura da indústria automobilística e de outros agentes atuantes no mercado (Moreira e Goldemberg, 1999). Diante do nível elevado das cotações de petróleo no mercado internacional, a expectativa da indústria é que essa participação se amplie ainda mais, iniciando uma onda de crescimento sem precedentes para o setor sucroalcooleiro. Em relação ao comportamento do preço do álcool hidratado, foram considerados na formação da série dados coletados no período de julho de 2000 a julho de 2007, com base semanal, referindo-se a negócios efetivados entre usinas e distribuidoras (preços ao produtor).

A soja e seus derivados constituem um dos produtos agrícolas mais comercializados em termos mundiais, dado que serve como principal insumo em diversos segmentos da cadeia Agroindustrial. Em relação ao comportamento de seus preços, foram considerados na formação da série dados coletados no período de agosto de 1997 a agosto de 2007, com base mensal, referindo-se a negociações no mercado físico de lotes (entre empresas).

A fonte básica para ambas as séries de preços referem-se aos indicadores de preços apurados pelo CEPEA (Centro de Estudos Avançados em Economia Aplicada) e divulgados em http://www.cepea.esalq.usp.br.

5.1 METODOLOGIA UTILIZADA PARA GERAÇÃO DE PREÇOS DO ALCÓOL HIDRATADO

O mercado nacional de álcool automotivo é composto pelas vendas de álcool anidro e de álcool hidratado. O consumo de álcool anidro está relacionado ao consumo de gasolina, visto que é misturado na proporção de 20% a 25% neste combustível. O álcool hidratado, por sua vez, é consumido pelos carros movidos puramente a álcool ou pelos carros *flex-fuel* que rodam com álcool e ou gasolina C, com mistura de 25% de etanol (Souza, 2006).

O álcool hidratado para uso em combustível é adquirido pelas distribuidoras e direcionado para os postos de revenda. A demanda pelo álcool hidratado depende de seu preço em relação à gasolina, uma vez que os consumidores são extremamente sensíveis ao diferencial de preços entre esses dois produtos, podendo migrar de um para outro, impactando a demanda do álcool. Em relação à produção, os preços maiores desse produto direcionam a cana à produção de álcool, sendo que, em sentido inverso, aumenta-se à produção de açúcar em detrimento da produção de álcool (Alves, 2002).

O levantamento de preços praticados, bem como os volumes equivalentes às transações, inclui as informações obtidas junto a uma amostra representativa dos segmentos que compõem o mercado do Estado de São Paulo, compreendendo usinas, destilarias, distribuidoras, e intermediários de vendas. As informações que compõem o sistema de informações referem-se a negócios efetivados, tendo como origem o Estado de São Paulo. O padrão do produto segue as especificações da Agência Nacional de Petróleo (ANP).

É apresentada na figura 5.1 a evolução dos preços do álcool combustível (álcool hidratado) no período de julho de 2000 a julho de 2007 na figura 5.1. Os preços coletados referem-se ao valor de faturamento diário (mas de divulgação semanal), referentes a negócios realizados no mercado físico, excluídos os impostos, praticados no mercado à vista ou em negociações a prazo com concessão de descontos financeiros, na relação Posto-Veículo-Usina (PVU) ou Posto-Veículo-Destilaria (PVD), expressos em reais. As transações baseadas em contratos entre usinas e clientes com

preços fixados para o ano-safra não são incluídas no cômputo da média. Incluem-se, porém, as transações em que os contratos estabelecem apenas o volume a ser entregue ao cliente, com preço a ser fixado por ocasião do faturamento.

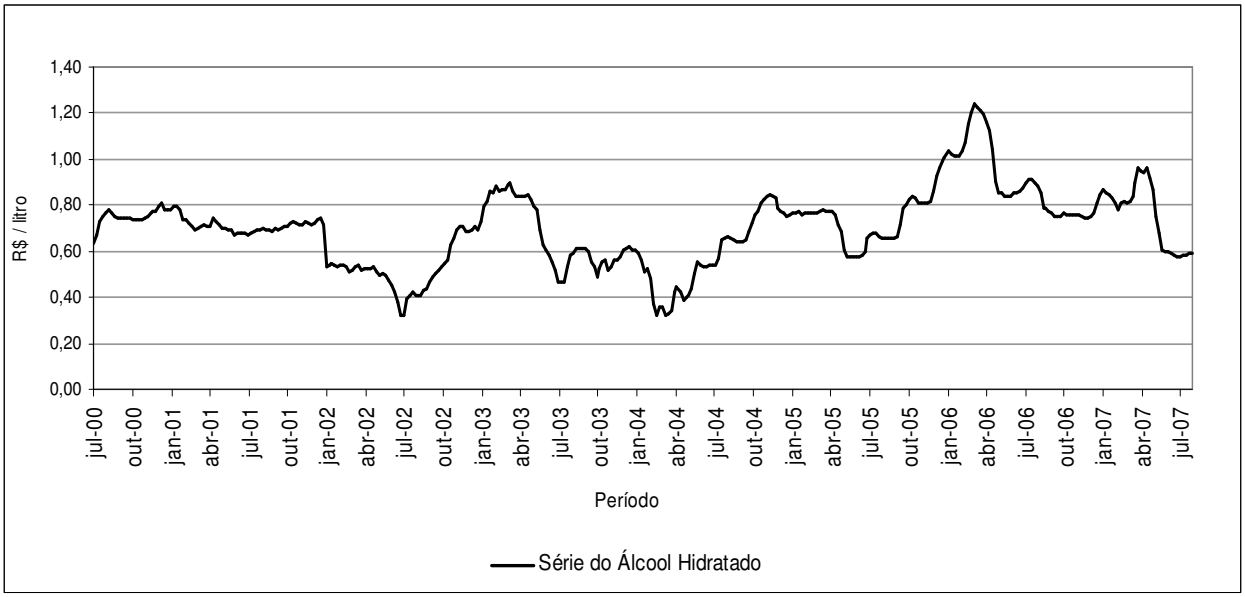

Figura 5.1: Relação de preços do Álcool Hidratado – Fonte: CEPEA (03/08/2007).

Dessa forma, este estudo tem como objetivo analisar a formação e transmissão de preços de um dos principais produtos do setor sucroalcooleiro, o álcool hidratado.

5.2 METODOLOGIA UTILIZADA PARA GERAÇÃO DA SÉRIE DE PREÇOS DO GRÃO DE SOJA

Segundo dados de 2007 do Departamento de Agricultura dos Estados Unidos da América (USDA, *United States Department of Agriculture*), os Estados Unidos, o Brasil e a Argentina ocupam expressiva participação no mercado internacional de soja e seus derivados no que se refere à produção e exportação de grãos desse produto. Portanto, somente esses três países respondem por quase 80% da produção mundial de grão de

soja.

Dada a importância econômica dessa *commodity*, são citados em Margarido *et al.* (2002) diversos estudos utilizando séries temporais na estimação do preço da soja no Brasil, como a análise da transmissão de preços do Porto de Rotterdam na Holanda sobre os preços de exportação deste subproduto no Brasil, Estados Unidos e Argentina. Aplicou-se sobre as séries temporais de preços o teste da raiz unitária do tipo Dickey-Fuller Aumentado, Modelos Auto-Regressivos Integrados de Médias Móveis (ARIMA, *AutoRegressive Integrated Moving Average*), Modelo de Função de Transferência, Teste de Co-integração de Engle-Granger e o Modelo de Correção de Erro (Freitas *et al.*, 2001).

Em Margarido e Sousa (1998), têm-se a análise de transmissão de preços do grão de soja entre a CBOT (*Chicago Board of Trade*) e os preços praticados em nível de Brasil e Paraná, mostrando que variações nas cotações da soja na CBOT são transmitida apenas parcialmente, e sem defasagem temporal para os preços em nível de produtor, tanto no Brasil, quanto no Paraná.

Pode-se ainda citar, um estudo ampliado sobre a soja na medição da elasticidade de transmissão de preços envolvendo a CBOT, preços do Porto de Rotterdam e os preços domésticos no Brasil e Argentina proposto em Machado e Margarido (2000). Os resultados obtidos mostram que variações nos preços do grão de soja em Rotterdam são transferidas mais intensamente e rapidamente para os preços domésticos dessa *commodity* no Brasil e Argentina, comparativamente àquelas variações originadas a partir da CBOT. Outro aspecto relevante é o fato de que os preços na Argentina são mais sensíveis às variações de preços internacionais do que no Brasil, refletindo possivelmente as características de cada mercado, dado que o Brasil possui um importante mercado doméstico consumidor de derivados de soja, enquanto que, na Argentina, quase toda produção destina-se exclusivamente ao mercado internacional, tornado seus preços domésticos mais sensíveis diante de variações nos preços externos.

Apesar da expressiva participação no mercado mundial, o Brasil é considerado mais como um país tomador de preços do que um país formador do preço da soja e seus derivados no mercado internacional (Azevedo, 2004).

Uma das maiores inquietações relacionadas ao agronegócio da soja são as influências do mercado consumidor, as questões relacionadas à produção de soja e a formação dos preços.

A formação do preço atualmente não se dá somente por oferta e demanda. A oscilação do preço da soja também se dá pela existência de forte especulação influenciada pela CBOT, além da flutuação do dólar ocasionada por variações cambiais.

Outro item a ser considerado na formação do preço recebido pelo exportador (FOB, *Free of Board*) é o prêmio de exportação, que deve ser somado à cotação da bolsa de Chicago. O prêmio pode ser positivo, representando um ágio ou negativo, representando um deságio sobre as cotações do produto na CBOT (Moraes, 2002). A participação do prêmio no preço recebido pelo exportador é significativa, chegando a elevá-lo em mais de 20% e reduzi-lo em até 5%. O prêmio possui um caráter sazonal, com maiores valores observados no período de entressafra e os menores no período de safra e exportação. Nesse período, os prêmios normalmente tornam-se negativos, isto é, o preço recebido pelo exportador fica abaixo das cotações da soja na CBOT. A elevação dos estoques nos principais produtores (Estados Unidos, Brasil e Argentina) influencia significativamente o valor do prêmio, reduzindo o valor pago da soja no porto de Paranaguá no Paraná.

Portanto, pode-se afirmar que o preço da soja no mercado brasileiro, como em outros países exportadores, é formado pelas seguintes variáveis: (i) cotação na Bolsa de Chicago; (ii) o prêmio; e (iii) a taxa cambial.

Na figura (5.2) pode-se analisar a evolução dos preços da soja na região do Paraná, no período de agosto de 1997 a agosto de 2007. Os preços coletados referem-se a negócios realizados no mercado físico (entre empresas) na região do estado do Paraná, com valores de faturamento diário, a serem descontados os impostos, praticados no mercado à vista ou em negociações a prazo, com concessão de descontos financeiros relativos a Notas Promissórias Rurais, expressos em reais.

Figura 5.2: Série de preços do Grão de Soja – Fonte: CEPEA (10/08/2007).

Dessa forma, este estudo tem como objetivo analisar a formação e transmissão de preços de um dos principais *commodities* do setor de agronegócios, a soja.

5.3 RESUMO DO CAPÍTULO

Como já mencionado, com a necessidade da comunidade em lidar com sistemas cada vez mais complexos e difíceis de serem modelados matematicamente, possibilitou a utilização de procedimentos de identificação e mais especificamente de previsão de séries temporais, pois permitem prever o que acontece a um processo.

A regra básica em estimação é utilizar o conhecimento *a priori* do processo e suas características físicas, quando se seleciona a estrutura e os parâmetros do modelo matemático.

A validação é o procedimento final da identificação de sistemas, sendo efetuada após a determinação da estrutura e estimação de parâmetros, e tem como objetivo verificar se o modelo matemático obtido é capaz de representar a dinâmica do sistema

em questão.

Neste capítulo foram apresentadas a descrição de duas séries temporais. As séries temporais foram as do preço do álcool hidratado (no período de julho de 2000 a julho de 2007) e a do preço do grão de soja (no período de agosto de 1997 à agosto de 2007).

Capítulo 6

RESULTADOS DE SIMULAÇÃO

Serão apresentados e discutidos neste capítulo, um estudo comparativo do desempenho dos algoritmos C-Médias Nebuloso (FCM), Gustafson-Kessel (GK), K-Médias Nebuloso (FKM), Gath-Geva Modificado (MGG) e Modelos de Misturas Gaussianas (GMM) com Maximização da Esperança (EM) esférica e elíptica no projeto de um sistema nebuloso, das seguintes séries temporais (i) série de preços do álcool hidratado e (ii) série de preços do grão de soja, visando a previsão de séries temporais em curtíssimo prazo, ou seja, um-passo-à-frente $(K = 1)$.

Para previsão foi escolhido um modelo matemático para a representação das séries temporais. A estrutura de previsão escolhida consiste de um modelo matemático com 2 entradas representadas por $[y(t-1), y(t-2)]$ e 1 saída (a saída prevista) representada por $\hat{y}(t)$.

Para cada estudo de caso foi construída uma tabela contendo os resultados das simulações e também figuras dos melhores resultados de cada algoritmo de agrupamento. Cada tabela contém todas as informações necessárias para análise, tais como: os coeficientes de correlação da etapa de estimação (R^2_{est}), validação (R^2_{val}) e a média harmônica (R^2_{mh}), além do valor de abertura das funções de pertinências Gaussianas (σ) e o respectivo número de regras.

Deve-se comentar que sempre é utilizada a metade das amostras K disponíveis de cada estudo de caso, para a fase de estimação dos parâmetos e a outra metade para a fase de validação. Para a série do álcool hidratado e a série do grão de soja, foram utilizadas respectivamente 183 e 1246 amostras tanto na fase de estimação, quanto na fase de validação. Com relação à abertura das funções de pertinência Gaussianas, elas não foram mantidas fixas. Para a série histórica do álcool hidratado e para a série do grão de soja, foram utilizados, respectivamente, os intervalos $[0{,}1; 100]$ e $[5; 1000]$.

6.1 ANÁLISE DOS RESULTADOS DE IDENTIFICAÇÃO

Após a obtenção do modelo de previsão baseado em modelo nebuloso de TS é realizada uma análise de desempenho do sistema para verificação se ele é capaz ou não de representar a dinâmica do sistema em questão.

O conhecimento da finalidade do modelo matemático se faz necessário, para poder julgar se ele incorpora ou não as características requeridas (Coelho e Alves, 2006).

Foram utilizadas regras de produção SE <condição> ENTAO <conclusão> do sistema nebuloso, estas com parte conseqüente de ordem zero, onde

$$\begin{aligned} &\text{SE} : z \text{ é } A_j(z), \\ &\text{ENTÃO} : \hat{y} = \frac{\sum_{j}^{K} A_{\text{j}}(z)\hat{\theta}_j}{\sum_{j}^{K} A_{\text{j}}(z)} \end{aligned} \tag{6.2}$$

onde

$$\hat{\theta} = [X^T X]^{-1} X^T y \tag{6.3}$$

6.1.1 Série do Álcool Hidratado

Conforme mencionado no capítulo 5, utiliza-se nesse trabalho a média harmônica do coeficiente de correlação múltipla das etapas de estimação e validação para avaliação do desempenho do modelo.

Na tabela 6.1 são apresentados de forma resumida os melhores resultados das simulações de previsão na fase de estimação e validação, usando de 2 a 6 funções de pertinência, variando-se o σ no intervalo $[0{,}1;\ 100]$.

Tabela 6.1: Resultados das Simulações do Modelo Nebuloso de TS – Série Álcool Hidratado.

σ	Método de Agrupamento	R^2 (estimação)	R^2 (validação)	R^2 (Média Harmônica)	Funções de pertinência Gaussianas para cada entrada do modelo nebuloso de TS	Número de Regras Utilizadas
0,1	C-Médias Nebuloso	0,931657	0,823922	0,874484	5	25
	K-Médias Nebuloso	0,929149	0,878570	0,903152	6	36
	Gustafson-Kessel	**0,952541**	**0,931023**	**0,941659**	6	36
	Gath-Geva Modificado	0,950216	0,809254	0,874088	6	36
	GMM-EM Esférico	0,931852	0,878721	0,904507	6	36
	GMM-EM Elíptico	0,929303	0,813511	0,867561	6	36
1	C-Médias Nebuloso	**0,963863**	0,974921	0,969361	5	25
	K-Médias Nebuloso	0,963787	0,969306	0,966538	5	25
	Gustafson-Kessel	0,963611	0,979968	0,971721	5	25
	Gath-Geva Modificado	0,963679	**0,981055**	**0,972289**	5	25
	GMM-EM Esférico	0,963796	0,967693	0,965741	5	25
	GMM-EM Elíptico	0,963599	0,979213	0,971343	4	16
2,5	C-Médias Nebuloso	**0,963782**	0,975939	0,969823	4	16
	K-Médias Nebuloso	0,963572	0,979393	0,971418	4	16
	Gustafson-Kessel	0,963554	**0,980095**	0,971754	5	25
	Gath-Geva Modificado	0,963606	0,979784	0,971628	4	16
	GMM-EM Esférico	0,963655	0,980033	**0,971775**	4	16
	GMM-EM Elíptico	0,963610	0,979797	0,971636	4	16
5	C-Médias Nebuloso	0,963610	0,979743	0,971609	4	16
	K-Médias Nebuloso	0,963602	0,979693	0,971581	4	16
	Gustafson-Kessel	0,963608	0,973737	0,971606	4	16
	Gath-Geva Modificado	0,963609	0,979772	0,971623	4	16
	GMM-EM Esférico	**0,963615**	0,979788	0,971634	4	16
	GMM-EM Elíptico	0,963603	**0,979826**	**0,971647**	3	9
10	C-Médias Nebuloso	0,958504	0,978489	0,968394	4	16
	K-Médias Nebuloso	0,961489	**0,980172**	0,970741	4	16
	Gustafson-Kessel	0,963180	0,977291	0,970184	3	9
	Gath-Geva Modificado	0,963213	0,980035	0,971552	4	16
	GMM-EM Esférico	0,957082	0,978598	0,967721	4	16
	GMM-EM Elíptico	**0,963592**	0,979829	**0,971642**	3	9
25	C-Médias Nebuloso	0,953256	0,971006	0,962049	3	9
	K-Médias Nebuloso	0,952358	**0,980335**	0,966144	4	16
	Gustafson-Kessel	0,960293	0,976993	0,968571	3	9
	Gath-Geva Modificado	**0,963573**	0,979721	**0,971580**	3	9
	GMM-EM Esférico	0,960461	0,979206	0,969743	6	36
	GMM-EM Elíptico	0,954942	0,979011	0,966827	3	9
50	C-Médias Nebuloso	0,931495	0,945582	0,938485	2	4
	K-Médias Nebuloso	0,931640	0,945711	0,938623	2	4
	Gustafson-Kessel	0,951976	0,972130	0,961948	5	25
	Gath-Geva Modificado	0,956429	0,975807	0,966021	4	16
	GMM-EM Esférico	0,931555	0,945636	0,938543	2	4
	GMM-EM Elíptico	**0,963468**	**0,979547**	**0,971441**	3	9
100	C-Médias Nebuloso	0,931495	0,945581	0,938485	2	4
	K-Médias Nebuloso	0,931640	0,945712	0,938623	2	4
	Gustafson-Kessel	0,931337	0,945439	0,938335	2	4
	Gath-Geva Modificado	0,932520	0,946501	0,939458	2	4
	GMM-EM Esférico	0,931556	0,945636	0,938543	2	4
	GMM-EM Elíptico	**0,945144**	**0,973544**	**0,959134**	3	9

Analisando-se os coeficientes de correlação (R^2), nenhum dos algoritmos apresentou valores menores que $0{,}9$ para a fase de estimação. No entanto, para a fase de validação apenas o algoritmo GK apresentou valor maiores que $0{,}9$ para $\sigma = 0{,}1$.

Um aspecto observado é que todos os algoritmos se mostraram eficientes para $\sigma \geq 1$, onde os melhores resultados de R^2 foram apresentados pelo algoritmo FCM (para a fase de estimação com $R^2{}_{est} = 0{,}963863$) e pelo algoritmo MGG (para a fase de validação com $R^2{}_{mh} = 0{,}981055$ e média harmônica igual a $0{,}972289$.

Nas figuras 6.1a e 6.1b são apresentados as saídas real e estimada, além do erro do melhor resultado de média harmônica $(R^2{}_{mh} = 0{,}972289)$, para o valor de abertura das funções de pertinências Gaussianas igual a 1 e suas 5 funções de pertinência de $y(t-1)$ e $y(t-2)$. Esse resultado foi obtido com a utilização do modelo nebuloso de TS e o método de agrupamento de Gath-Geva Modificado. Porém, esse resultado apresentou uma alta complexidade computacional em virtude das 25 regras utilizadas para a entrada do modelo nebuloso, que são apresentadas na Tabela A.1 do Anexo A.

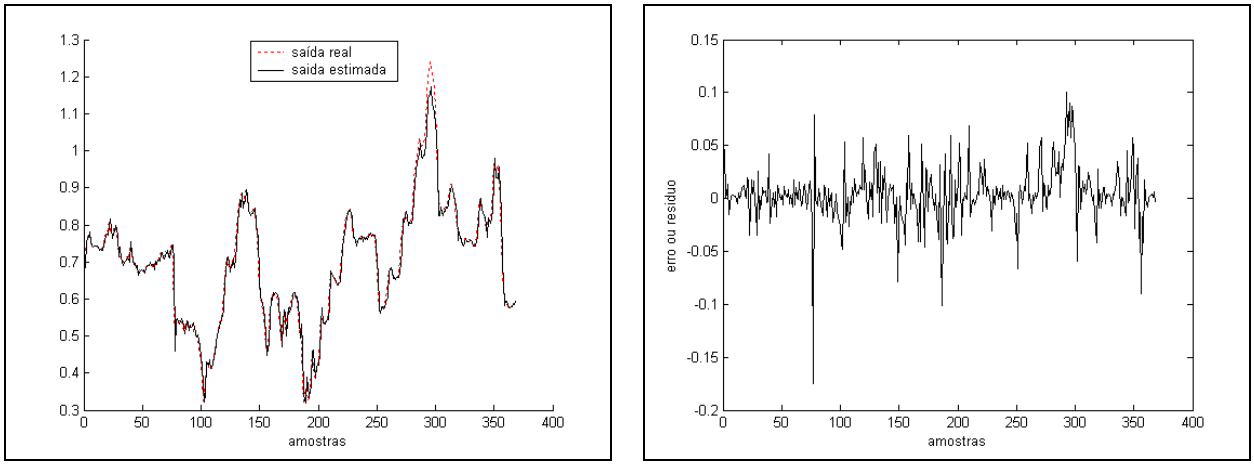

Figura 6.1a: Saída real e estimada e o erro do melhor resultado obtido usando modelo nebuloso de TS e método de agrupamento de Gath-Geva Modificado (MGG).

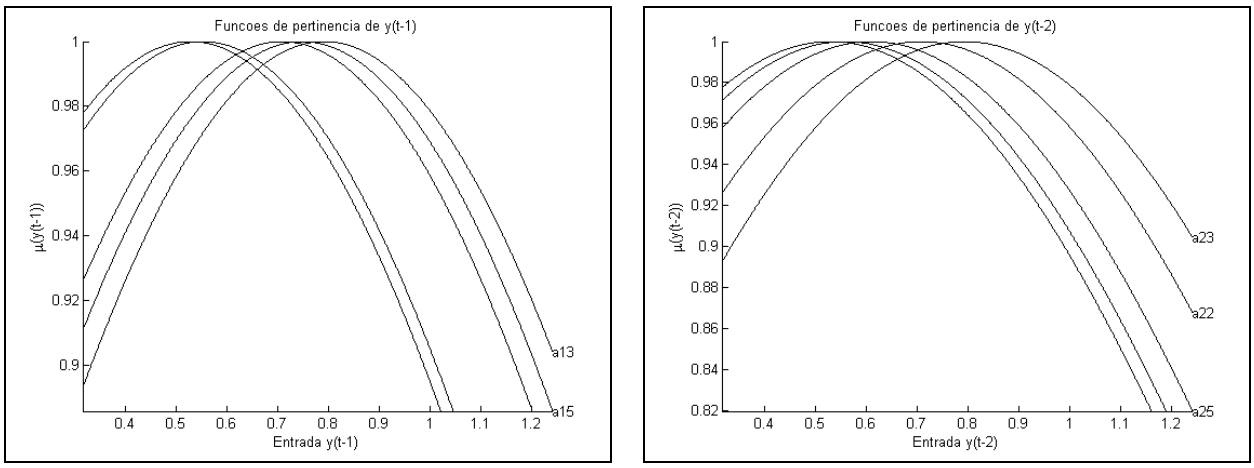

Figura 6.1b: Funções de pertinência Gaussianas de $y(t-1)$ e $y(t-2)$.

6.1.2 Série da Soja

Na tabela 6.2 são apresentados de forma resumida os melhores resultados das simulações de previsão na fase de estimação e validação, usando de 2 a 6 funções de pertinência, variando σ no intervalo $[5;1000]$.

Analisando-se os coeficientes de correlação (R^2), nenhum dos algoritmos apresentou valores menores que $0{,}9$ para a fase de estimação. No entanto, um aspecto observado é que todos os algoritmos não foram eficientes para $\sigma \leq 5$, apresentando valores menores que $0{,}9$ para a fase de estimação e média harmônica.

No entanto, todos os algoritmos se mostraram eficientes para $\sigma > 25$, onde os melhores resultados de R^2 foram apresentados pelo algoritmo MGG (para a fase de estimação com $R^2_{est} = 0{,}996740$) e pelo algoritmo de Mistura Gaussiana com Maximização da Esperança Elíptica (para a fase de validação com $R^2_{mh} = 0{,}996065$ e média harmônica igual a $0{,}996398$).

Tabela 6.2: Resultados das Simulações do Modelo Nebuloso de TS – Série Soja.

σ	Método de Agrupamento	R^2 (estimação)	R^2 (validação)	R^2 (Média Harmônica)	Funções de pertinência Gaussianas para cada entrada do modelo nebuloso de TS	Número de Regras Utilizadas
5	C-Médias Nebuloso	**0,995019**	0,516861	0,680328	6	36
	K-Médias Nebuloso	0,993549	0,250508	0,400129	5	25
	Gustafson-Kessel	0.988951	0,000000	0,000000	6	36
	Gath-Geva Modificado	0,994152	**0,618295**	**0,762418**	5	25
	GMM-EM Esférico	0,994166	0,273885	0,429458	6	36
	GMM-EM Elíptico	0,994522	0,485653	0,652615	5	25
10	C-Médias Nebuloso	0,994846	0,921481	0,956759	5	25
	K-Médias Nebuloso	0,994861	0,861905	0,923623	5	25
	Gustafson-Kessel	0,994806	0,744383	0,851565	4	16
	Gath-Geva Modificado	0,994829	0,964705	0,979536	5	25
	GMM-EM Esférico	0,994864	0,861198	0,923218	5	25
	GMM-EM Elíptico	**0,996632**	**0,992335**	**0,994478**	6	36
25	C-Médias Nebuloso	**0,994816**	0,990571	0,992689	3	9
	K-Médias Nebuloso	0,994815	0,991408	0,993109	3	9
	Gustafson-Kessel	0,994803	**0,993157**	**0,993979**	2	4
	Gath-Geva Modificado	0,994815	0,992540	0,993676	2	4
	GMM-EM Esférico	0,994816	0,990871	0,992839	3	9
	GMM-EM Elíptico	0,994818	0,992993	0,993905	3	9
50	C-Médias Nebuloso	0,994805	0,993317	0,994060	2	4
	K-Médias Nebuloso	0,994919	0,915713	0,953674	4	16
	Gustafson-Kessel	0,996740	0,991567	0,994147	5	25
	Gath-Geva Modificado	**0,996740**	0,991604	0,994165	5	25
	GMM-EM Esférico	0,994805	0,993318	0,994061	2	4
	GMM-EM Elíptico	0,996732	**0,996065**	**0,996398**	4	16
100	C-Médias Nebuloso	0,994804	0,993212	0,994007	2	4
	K-Médias Nebuloso	0,994804	0,993212	0,994007	2	4
	Gustafson-Kessel	0,996717	0,994010	0,995362	4	16
	Gath-Geva Modificado	**0,996737**	**0,995155**	**0,995945**	4	16
	GMM-EM Esférico	0,994804	0,993212	0,994007	2	4
	GMM-EM Elíptico	0,996737	0,995149	0,995943	4	16
250	C-Médias Nebuloso	0,994804	0,993203	0,994003	2	4
	K-Médias Nebuloso	0,994804	0,993203	0,994003	2	4
	Gustafson-Kessel	0,994802	0,993201	0,994001	2	4
	Gath-Geva Modificado	**0,996728**	**0,995839**	**0,996283**	4	16
	GMM-EM Esférico	0,994804	0,993203	0,994003	2	4
	GMM-EM Elíptico	0,996678	0,995801	0,996239	4	16
500	C-Médias Nebuloso	0,994804	0,993203	0,994003	2	4
	K-Médias Nebuloso	0,994803	0,993203	0,994002	2	4
	Gustafson-Kessel	0,994802	0,993201	0,994001	2	4
	Gath-Geva Modificado	**0,994812**	**0,993215**	**0,994013**	2	4
	GMM-EM Esférico	0,994804	0,993203	0,994003	2	4
	GMM-EM Elíptico	**0,994812**	**0,993215**	**0,994013**	2	4
1000	C-Médias Nebuloso	0,994804	0,993203	0,994003	2	4
	K-Médias Nebuloso	0,994804	0,993203	0,994003	2	4
	Gustafson-Kessel	0,994802	0,993201	0,994001	2	4
	Gath-Geva Modificado	**0,994812**	0,993215	0,994013	2	4
	GMM-EM Esférico	0,994742	**0,993307**	**0,994024**	3	9
	GMM-EM Elíptico	0,994811	0,993215	0,994013	2	4

Nas figuras 6.2a e 6.2b são apresentados as saídas real e estimada, além do erro do melhor resultado de média harmônica $\left(R^2_{mh} = 0{,}996398\right)$, para o valor de abertura das funções de pertinências Gaussianas igual a 50 e suas quatro funções de pertinência de $y(t-1)$ e $y(t-2)$.

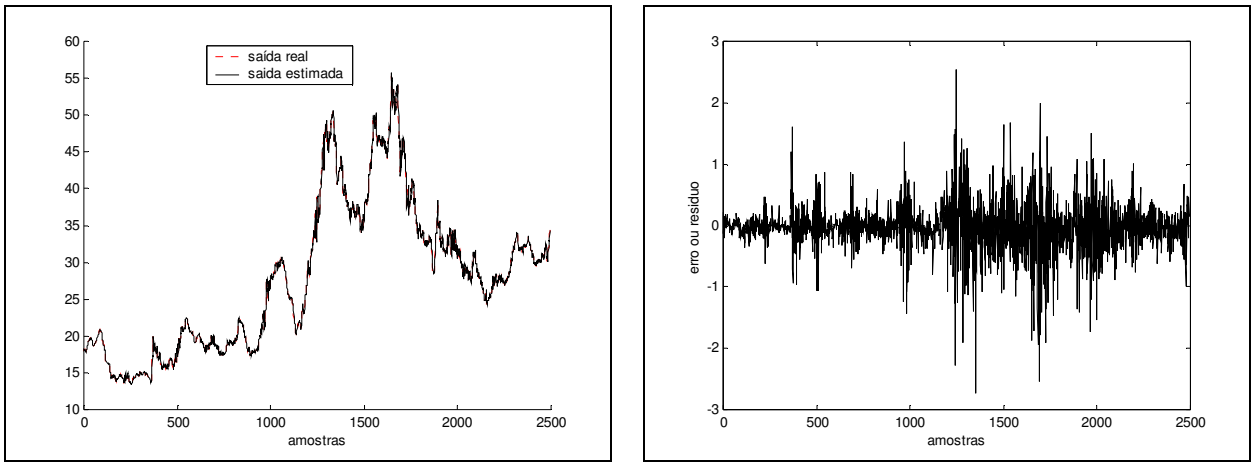

Figura 6.2a: Saída real e estimada e o erro do melhor resultado obtido usando modelo nebuloso de TS e método de agrupamento de Misturas Gaussianas com Maximização da Esperança Elíptica (GMM-EM).

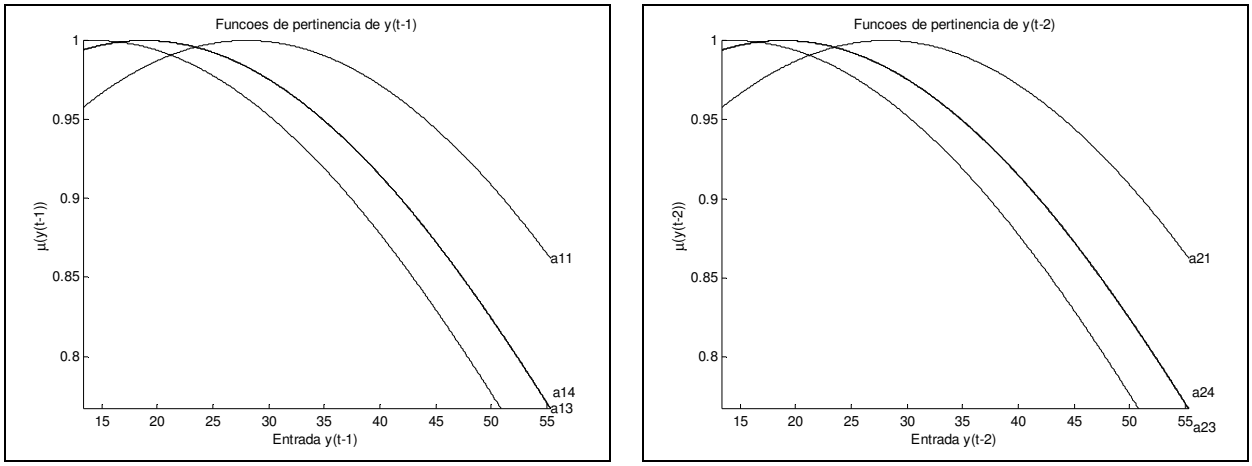

Figura 6.2b: Funções de pertinência Gaussianas de $y(t-1)$ e $y(t-2)$.

Esse resultado foi obtido com a utilização do modelo nebuloso de TS e o método de agrupamento de Misturas Gaussianas com Maximização da Esperança Elíptica. Porém, esse resultado apresentou uma alta complexidade computacional em virtude das 16 regras utilizadas para a entrada do modelo nebuloso, que são apresentadas na Tabela A.2 do Anexo A.

6.2 RESUMO DO CAPÍTULO

Foram utilizadas de 2 a 6 funções de pertinência nos testes de previsão com o modelo nebuloso de Takagi-Sugeno, pois com apenas 1 função de pertinência as previsões ficaram com valores de R^2 em torno de $0{,}40$, este um resultado insatisfatório. para fins de previsão Em outro extremo, observa-se que para os casos testados com mais de 6 funções de pertinência para cada entrada do modelo nebuloso não foi obtida uma melhora significativa nos valores de R^2.

Com relação ao σ, conseguiu-se desempenhos satisfatórios para valores compreendidos respectivamente entre 1 e 100 para a série do álcool, e entre 10 e 1000 para a série da soja. Neste contexto, foram realizados vários testes com diferentes valores de σ. Para valores de σ abaixo desses intervalos, os valores de R^2 foram considerados insatisfatórios e para valores de σ acima desses intervalos, os valores de R^2 pouco se alteraram.

Deve-se enfatizar que os valores de σ para as simulações foram obtidos por heurística de tentativa e erro. Neste contexto, melhores resultados aos que foram apresentados nas tabelas 6.1 e 6.2, poderiam ter sido obtidos com a inclusão de um método de otimização para ajuste do σ. No entanto, este não foi o foco deste trabalho.

Na série de preços do álcool hidratado, observaram-se os melhores resultados com a utilização do método de agrupamento de Gath-Geva modificado, representado na figura 6.1. Para $\sigma = 1$, os valores dos coeficientes de correlação de estimação e validação foram respectivamente $0{,}963679$ e $0{,}981055$, e a média harmônica igual a $0{,}972289$.

Na série de preços do grão de soja, observaram-se os melhores resultados com a utilização do método de agrupamento C-Médias Nebuloso, representado na figura 6.2. Para $\sigma = 50$, os valores dos coeficientes de correlação de estimação e validação foram respectivamente $0{,}996740$ e $0{,}996065$, e a média harmônica igual a $0{,}996398$.

Capítulo 7

COMENTÁRIOS GERAIS E TRABALHOS FUTUROS

Este capítulo final apresenta os comentários gerais, resume os resultados e as contribuições deste trabalho, bem como sugere possíveis trabalhos futuros.

7.1 COMENTÁRIOS GERAIS

Os algoritmos de agrupamento de dados são usados não somente para organizar, categorizar ou descobrir a relevância dos dados, mas também são úteis à compressão dos dados e a construção de modelos matemáticos. Dessa forma, algumas considerações devem ser realizadas para cada algoritmo de agrupamento de dados avaliado neste trabalho para o projeto do modelo nebuloso de TS.

O algoritmo FCM (*Fuzzy C-Means*) é sensível às escolhas iniciais dos grupos, precisando de uma inicialização adequada para assegurar a obtenção de resultados corretos. A utilização de algoritmos para realizar essas escolhas pode resolver o problema, assegurando assim uma correta convergência dos centros do grupo e a redução do número de iterações necessárias.

Com relação ao algoritmo GK (Gustafson-Kessel), este introduz uma métrica corrigida (distância de Mahalanobis) para permitir que os grupos possam ser alongados (elipsóides). Esse procedimento é um processo interativo, na qual os centros dos grupos são movidos no espaço (entrada-saída) até que algum critério de convergência seja encontrado.

Uma característica do algoritmo FKM (*Fuzzy K-Means*), é que a localização inicial dos centros pode variar permitindo estabelecer outras condições iniciais para que o algoritmo possa melhorar seu desempenho. Outro aspecto a ser comentado é que o

FKM faz com que cada dado do conjunto pertença a um grupo. Porém, essa característica pode se tornar uma desvantagem, pois mesmo o dado não estando tão próximo ao centro, ele pertencerá a este, contribuindo na redução de sua eficiência. Outra desvantagem é que o número de grupos deve ser escolhido antes de iniciar o algoritmo.

Já o algoritmo GG (Gath-Geva) é um método pouco robusto, pois relaciona a distância do centro a uma distribuição de probabilidades. Devido sua função de pertinência decrescer lentamente a partir dos respectivos centros, o algoritmo converge para um mínimo local próximo. Para contornar este problema, o algoritmo Gath-Geva é modificado (Modified Gath-Geva), baseado na maximização da esperança de modelos Gaussianos, sendo utilizado para estimar a densidade de probabilidade de um conjunto de dados.

Em relação aos resultados, pode-se realizar uma análise direta dos mesmos. Para se obter uma melhor resposta da previsão de séries temporais, foram apresentados os índices R^2 (nas etapas de estimação e validação) e R^2_{mh}.

Por exemplo, na análise da série do álcool utilizam-se 25 regras (com 5 funções de pertinência) para se obter R^2 igual a 0,972289. Para a mesma série, utilizam-se 04 regras (com duas funções de pertinência) para se obter R^2 igual a 0,939482. Considerando-se que o número de regras utilizadas determina a complexidade computacional do modelo e que valores de R^2 no intervalo $[0{,}9; 1]$ são considerados suficientes para aplicações práticas, pode-se apresentar R^2 igual a 0,939482 como melhor resultado, devido sua quantidade menor de regras em relação a R^2 igual a 0,972289.

Outro importante aspecto é a dificuldade em analisar qual dos algoritmos de agrupamento de dados obteve um melhor resultado, pois todos obtiveram resultados próximos. Para determinar o melhor algoritmo, seria necessário a realização de uma rigorosa análise estatística dos resultados, bem como a adoção e estudo detalhado de um procedimento de otimização para os valores de σ.

Pôde-se também observar que os valores de R^2_{est} e R^2_{val} sempre foram distintos, devido a diferença entre a fase de estimação e validação.

Determinando-se de que forma deseja-se obter os resultados e qual o modelo mais adequado, este pode ser utilizado na previsão do comportamento do sistema a curtíssimo prazo (um-passo-à-frente).

7.2 TRABALHOS FUTUROS

Esta dissertação teve por objetivo apresentar um procedimento de otimização de um modelo nebuloso do tipo TS usando diversos métodos de agrupamento de dados, para identificação dos antecedentes (premissa das regras do modelo nebuloso) aliado ao método dos mínimos quadrados para o ajuste dos conseqüentes das regras nebulosas, para previsão um-passo-à-frente de séries temporais na área de agronegócios. As séries históricas testadas foram: (i) série histórica de preços do álcool hidratado e (ii) série histórica dos preços dos grãos da soja.

Os objetivos de estudar e analisar os resultados de simulação em identificação, tanto da fase de estimação quanto validação dos modelos foram atingidos, pois os resultados obtidos foram satisfatórios.

Em síntese, a contribuição desta dissertação foi a de validar em dois estudos de casos o uso do modelo nebuloso Takagi-Sugeno e também de métodos de agrupamento.

Para futura pesquisa, a utilização de outros algoritmos para otimizar os centros o número de abertura das funções de pertinência Gaussianas (σ) para detecção do valor otimizado de cada abordagem de agrupamentos de dados usando métodos de computação evolutiva e inteligência de enxames (ou inteligência coletiva).

REFERÊNCIAS BIBLIOGRÁFICAS

ABELÉM, A.J.G. **Redes Neurais Artificiais na Previsão de Séries Temporais.** Dissertação de Mestrado. Departamento de Engenharia Elétrica da Pontifícia Universidade Católica do Rio de Janeiro, PUC-Rio. Rio de Janeiro, RJ. 1994. 100p.

ABONYI, J. ***Fuzzy Model Identification for Control.*** Birkhäuser, Boston, USA. 2003. 290p.

ABONYI, J.; BABUŠKA; R.; SZEIFERT, F. ***Modified Gath-Geva Fuzzy Clustering for Identification of Takagi-Sugeno Fuzzy Models.*** IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics, vol. 32, n. 5, 2002. p. 612-621.

ABONYI, J.; FEIL, B.; NEMETH, S.; ARVA, P. ***Modified Gath-Geva Clustering for Fuzzy Segmentation of Multivariate Time Series.*** Fuzzy Sets and Systems, vol. 149, n. 1, 2005. p. 39-56.

AGUIRRE, L.A. **Introdução à Identificação de Sistemas: Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais.** Universidade Federal de Minas Gerais, UFMG. Belo Horizonte, MG, Brasil. 2000. 554p.

AHMAD, S.M.; SUTTON, R. ***Linear, Nonlinear Modeling and AUV Retrieval.*** Technical Report, Department of Mechanical and Marine Engineering, University of Plymouth, United Kingdom, 2000.

ALMEIDA, F.M. **Identificação Multivariável de um Processo de Incineração de Resíduos Líquidos Utilizando Modelos Nebulosos Takagi-Sugeno.** Dissertação de Mestrado. Faculdade de Engenharia Elétrica e de Computação. Universidade Estadual de Campinas, UNICAMP. Campinas, SP. 2005. 138p.

ALVES, L.R.A. **Transmissão de Preços entre Produtos do Setor Sucroalcooleiro do Estado de São Paulo**. Dissertação de Mestrado. Escola Superior de Agricultura Luiz de Queiroz. Universidade de São Paulo, USP. Piracicaba, SP. 2002. 123p.

ANGELOV, P.P.; FILEV, D.P. ***An Approach to Online Identification of Takagi-Sugeno Fuzzy Models.*** IEEE Transactions on Systems, Man and Cynernetics - Part B: Cybernetics, vol. 34, n. 1, 2004. p. 484-498.

AZEVEDO, M.S. **Preço Internacional e Produção de Soja no Brasil de 1995 a 2003.** Relatório Técnico. Caderno de Ciências Econômicas. Uni-Anhaguera. Goiânia, GO. 2004. 7p.

BABUŠKA, R. ***Fuzzy Modeling for Control.*** Kluwer Academic Publishers, Norwell, USA, 1998. 260p.

BILLINGS, S. ***Identification of Nonlinear Systems – A Survey.*** IEEE Proceedings, Control Theory and Applications, Part D, vol. 127, n. 6, 1980. p.272-285.

BOLLERSLEV, T. ***Generalized Autoregressive Conditional Heteroscedasticity*****.** Journal of Econometrics, vol. 31, 1986. p. 07-327.

BONVENTI JR., W. **Aprendizado Nebuloso Híbrido e Incremental para Classificar *Pixels* por Cores.** Tese de Doutorado. Escola Politécnica, Universidade de São Paulo, USP. São Paulo, SP, Brasil. 2005. 177p.

BONVENTI JR. W.; COSTA, A.H.R. **Sistema Semi-Automático de Detecção de Pele por Agrupamentos Nebulosos.** VI Simpósio Brasileiro de Automação Inteligente. Bauru, SP, Brasil. 2003. 6p.

BOX, G.E.P.; JENKINS, G.M. ***Time Series Analysis, Forecasting and Control.*** San Francisco, CA: Holden Day, 1970. 575p.

BURLAMAQUI, C.S.; LIANG, Y.C. **Teoria dos Sistemas Nebulosos Aplicados a Sistemas de Informação Geográfica para Definição de Roteiro.** IV Simpósio Brasileiro de Geo-Informática. Caxambú, MG, Brasil. 2002. p. 9-16.

CASSINI, C.C.S.; AGUIRRE, L.A. **Uma Introdução de Sistemas Não-Lineares.** Laboratório de Modelagem, Análise e Controle de Sistemas Não-Lineares. Relatório Técnico, Departamento de Engenharia Elétrica da Universidade Federal de Minas Gerais, UFMG. Belo Horizonte, MG. 1999. 23p.

CASTANHO, M.J.P. **Construção e Avaliação de um Modelo Matemático para Predizer a Evolução do Câncer de Próstata e Descrever seu Crescimento Utilizando a Teoria dos Conjuntos Fuzzy.** Tese de Doutorado. Faculdade de Engenharia Elétrica e de Computação. Universidade Estadual de Campinas, UNICAMP. Campinas, SP. 2005. 148p.

CIRIGLIANO, R.J.R. **Identificação de Locutor: Otimização do Número de Componentes Gaussianas.** Dissertação de Mestrado. Universidade Federal do Rio de Janeiro, UFRJ. Rio de Janeiro, RJ. 2007. 72p.

CHEN, T.L.; GHENG, C.H.; TEOH, H.J. ***Fuzzy Time Series based on Fibonacci Sequence for Stock Price Forecasting.*** Physica A: Statistical Mechanics and its Applications, vol. 380, 2007. p. 377-390.

CHENG, C.T.; OU, C.P.; CHAU, K.W. ***Combining a Fuzzy Optimal Model with a Genetic Algorithm to Solve Multi-Objective Rainfall-Runoff Model Calibration.*** Journal of Hydrology, vol. 268, n. 1-4, 2002. p. 72-86.

COELHO, L.S. **Identificação e Controle de Processos Multivariáveis via Metodologias Avançadas e Inteligência Computacional.** Tese de Doutorado. Centro Tecnológico da Universidade Federal de Santa Catarina, UFSC. Florianópolis, SC. 2000. 342p.

COELHO, L.S. **Fundamentos, Potencialidades e Aplicações de Algoritmos Evolutivos.** São Carlos, SP, Brasil. 2003. 111p.

COELHO, L.S.; ALVES, L. **Identificação Não-Linear de um Trocador de Calor Baseada em um Modelo Nebuloso Interpolativo.** Proceedings of the ENCIT 2006, número do artigo: CIT 06-0321, 2006, p. 5-8.

DELGADO, M.R.B.S. **Projeto Automático de Sistemas Nebulosos: Uma Abordagem Co-Evolutiva.** Tese de Doutorado. Faculdade de Engenharia Elétrica e da Computação. Universidade Estadual de Campinas, UNICAMP. Campinas, SP. 2002. 204p.

DOMINGUES, H.H. **Espaços Métricos e Introdução à Topologia**. São Paulo, SP: Atual, 1982. 184p.

DÖRING, C.; LESOT, M.J.; KRUSE, R. ***Data Analysis with Fuzzy Clustering Methods.*** Computational Statistics & Data Analysis, vol. 51, n. 1, 2006. p. 192-214.

DOTE, Y.; OVASKA, S.J. ***Industrial Applications of Soft Computing: A Review.*** Proceedings of the IEEE, vol. 89, n. 9, 2001. p.1243-1265.

DU, H.; ZHANG, N. ***Application of Evolving Takagi–Sugeno Fuzzy Model to Nonlinear System Identification.*** Applied Soft Computing, vol. 8, n. 1, p. 675-585, 2008.

FAVARO, G.M. **Dinâmicas Autoregressivas em Econofísica.** Dissertação de Mestrado. Instituto de Física de São Carlos. Universidade de São Paulo, USP. São Carlos, SP. 2007. 91p.

FREITAS, S. M.; MARGARIDO, M.A.; BARBOSA, M.Z.; FRANCA, T.J.F. **Análise da Dinâmica de Transmissão de Preços no Mercado Internacional de Farelo de Soja, 1990-99.** Agricultura em São Paulo, vol. 48, n. 1, 2001. p. 1-20.

GAO, Y.; JOO, M. E. ***Online Adaptive Fuzzy Neural Identification and Control of a Class of MIMO Nolinear Systems.*** IEEE Transactions on Fuzzy Systems, vol. 11, n. 4, 2003. p. 462-477.

GATH, I.; GEVA, A.B. ***"Unsupervised Optimal Fuzzy Clustering,"*** IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, n. 7, 1989. p. 211-218.

GOOIJER, J.G.; HYNDMAN, R.J. ***25 Years of Time Series Forecasting.*** International Journal of Forecasting, vol. 22, n. 3, 2006. p. 443-473.

GORP, J.V. **Nonlinear Identification with Neural Network and Fuzzy Logic.** Tese de Doutorado. Department of Fundamental Electricity and Instrumentation. Vrije Universiteit Brussel, Brussel, Bélgica, 2000. 280p.

GREBLICK, W.; PAWLAK, M. ***Recursive Non Parametric Identification of Hammerstein Systems.*** Journal of the Franklin Institute, vol. 326, n. 4, 1989. p. 461-481.

GUERRA, F.A. **Análise de Métodos de Agrupamento para o Treinamento de Redes Neurais de Base Radial Aplicadas à Identificação de Sistemas.** Dissertação de Mestrado. Programa de Pós-Graduação em Engenharia de Produção e Sistemas. Pontifícia Universidade Católica do Paraná, PUC. Curitiba, PR. 2006. 149p.

HAMMOUDA, K.; KARRAY, F. ***A Comparative Study of Data Clustering Techniques.*** University of Waterloo, Ontario, Canadá. 2000. p. 1-21.

HÄRDLE, W.;HLÁVKA,Z.;KLINKE,S. ***Xplore Application Guide.*** MDTech, Springer: 2003. 553p.

JAIN, A.K.; MURTY, M.N.; FLYNN, P.J. ***Data Clustering: a Review.*** ACM Computing Surveys, vol. 31, n. 3, 1999. p. 264-323.

KRISHNAPURAM, R.; KIM, J. ***A Note on the Gustafson-Kessel and Adaptive Fuzzy Clustering Algorithms.*** IEEE Transactions on Fuzzy Systems, vol. 7, n. 4, 1999. p. 453-461.

LANDMANN, R. **Um Modelo Heurístico para a Programação da Produção em Fundições com Utilização da Lógica Fuzzy.** Tese de Doutorado. Programa de Pós-Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina, UFSC. Florianópolis, SC, Brasil. 2005. 207p.

LIMA, C.A.M. **Comitê de Máquinas: Uma Abordagem Unificada Empregando Máquinas de Vetores-Suporte.** Tese de Doutorado. Faculdade de Engenharia Elétrica e de Computação. Universidade Estadual de Campinas, UNICAMP. Campinas, SP. 2004. 378p.

LIN, C.T.; CHEN, C.B.; WU, W.H. ***Fuzzy Clustering Algorithm for Latent Class Model.*** Statistics and Computing, vol. 14, n. 4, 2004. pp.299-310.

LJUNG, L.; GLAD, T. ***Modeling of Dynamic Systems.*** Prentice-Hall, Uper Saddle River, NJ, USA. 1994. 361p.

LUCHETTA, A.; MANETTI, S. A. ***Real Time Hydrological Forecasting System Using a Fuzzy Clustering Approach.*** Computers & Geosciences, vol. 29, n. 9, 2003. p. 1111-1117.

LUNA, J.E.O. **Algoritmos EM para Apredizagem de Redes Bayesianas a Partir de Dados Incompletos.** Dissertação de Mestrado. Departamento de Computação e Estatística. Universidade Federal de Mato Grosso do Sul. Campo Grande, MS. 2004. 132p.

MACHADO, E.R.M.D. **Modelagem e Controle de Sistemas Fuzzy Takagi-Sugeno.** Tese de Doutorado. Faculdade de Engenharia, Universidade Estadual Paulista, UNESP. Ilha Solteira, SP. 2003. 209p.

MACHADO, M.A.S.; FILHO, P.S.B.; MEDEIROS, V.Z. **Um Sistema de Inferência Nebuloso para Apoio à Tomada de Decisão do Analista de Crédito de Empresas de Crédito Pessoal.** RESI – Revista Eletrônica de Sistemas de Informação. Edição 7, ano V, n. 1. 2006. p. 1-25.

MACHADO, E. L.; MARGARIDO, M. A. ***Seasonal Price Transmission in Soybean International Market: the Case of Brazil and Argentina.*** Anais do XXVIII Congresso Brasileiro de Economia e Sociologia Rural. Rio de Janeiro, RJ. 2000. 16p.

MAGALHÃES, M.H. **Redes Neurais, Metodologias de Agrupamento e Combinação de Previsores Aplicados à Previsão de Vazões Naturais.** Dissertação de Mestrado. Departamento de Engenharia de Computação e Automação Industrial. Universidade Estadual de Campinas, UNICAMP. Campinas, SP. 2004. 123p.

MAMDANI, E.H.; ASSILIAN, S. ***An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller.*** International Journal Man-Machine Studies, vol. 7, n. 1, 1975. p. 1-15.

MARGARIDO, M. A.; SOUSA, E. L. L. **Formação de Preços da Soja no Brasil.** Agricultura em São Paulo, vol. 45, n. 2, 1998. p. 52-61.

MARGARIDO, M. A.; FERNANDES, J.M.; TUROLLA, F.A. **Análise da Formação de Preços no Mercado Internacional de Soja: o Caso do Brasil.** Agricultura em São Paulo, vol. 47, n. 2, 2002. p. 71-85.

MARTIN, C. **Aplicação de Redes Neurais para Prognóstico com Base em Séries Temporais.** Dissertação de Mestrado. Programa de Pós-Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina, UFSC. Florianópolis, SC. 2000. 87p.

MELIN, P.; MANCILLA, A.; LOPEZ, M.; MENDONZA, O. ***A hybrid modular neural network architecture with fuzzy Sugeno integration for time series forecasting***. Applied Soft Computing, vol. 7, n. 4, 2007. p. 1217-1226.

MOON, T. K. ***The Expectation – Maximization Algorithm.*** IEEE Signal Processing Magazine, vol. 13, n. 6, 1996. p. 47-60.

MORAES, M. **Prêmio de Exportação da Soja Brasileira.** Dissertação de Mestrado. Escola Superior de Agricultura Luiz de Queiroz. Piracicaba. Universidade de São Paulo, USP. São Paulo, SP. 2002. 106p.

MOREIRA, J.R.; GOLDEMBERG, J. ***The Alcohol Program***. Energy Policy, vol. 27, n. 4, 1999. p. 229-245.

MORETTIN, P.A.; TOLOI, C.M. **Séries Temporais.** 2ª edição, Atual: São Paulo, SP, Brasil. 1987. 136p.

MUELLER, A. **Uma Aplicação de Redes Neurais Artificiais na Previsão do Mercado Acionário.** Dissertação de Mestrado. Centro Tecnológico da Universidade Federal de Santa Catarina, UFSC. Florianópolis, SC. 1996. 103p.

NEPOMUCENO, E.G. **Identificação Multiobjetivo de Sistemas Não-Lineares.** Dissertação de Mestrado. Departamento de Engenharia Elétrica. Universidade Federal de Minas Gerais, UFMG. Belo Horizonte, MG. 2002. 200p.

PAI, P.F. ***Hybrid Ellipsoidal Fuzzy Systems in Forecasting Regional Electricity Lodas.*** Energy Conversation and Management, vol. 47, n. 15-16. 2006. p. 2283-2289.

PASSINO, K.M.; YURKOVICH, S. ***Fuzzy Control.*** Addison-Wesley, CA, USA. 1998. 522p.

PEDRYCZ, W.; GOMIDE, F. ***An Introduction to Fuzzy Sets: Analysis and Design.*** Cambridge, UK: MIT Press, 1998. 465p.

PINA, J.M.; LIMA, P.U. ***A Glass Furnace Operation System using Fuzzy Modeling and Genetic Algorithms for Performance Optimization.*** Engineering Applications of Artificial Intelligence, vol.16, n. 7-8. 2003. p. 681-690.

PUCCIARELLI, A.J. **Modelagem de Séries Temporais Discretas Utilizando Modelo Nebuloso Takagi-Sugeno.** Dissertação de Mestrado. Faculdade de Engenharia Elétrica e de Computação. Universidade Estadual de Campinas, UNICAMP. Campinas, SP. 2005. 130p.

PURCOTE, A. **Previsão de Demanda Utilizando a Metodologia Box & Jenkins de Séries Temporias.** Monografia de Conclusão de Curso. Universidade Federal do Paraná, UFPR. Curitiba, PR. 2006. 76p.

RENÉ, J. ***Fuzzy Logic in Control.*** Delft, The Netherlands. 1995. 322p.

RENNERS, I. ***Data-Driven System Identification via Evolutionary Retrieval of Takagi-Sugeno Fuzzy Models.*** Tese de Doutorado. Universidade de Magdeburg. Magdeburg, Alemanha. 2004. 183p.

REYES, C.A.P. ***Coevolutionary Fuzzy Modeling.*** Tese de Doutorado. École Polytechnique Fédérale de Lausanne. Lausanne, Switzland. 2002. 162p.

RIBEIRO, J.F.F.; MEGUELATI, S. **Organização de um Sistema de Produção em Células de Fabricação.** Gestão & Produção, vol. 9, n. 1, 2002. p. 62-67.

ROCHA, F.G. **Contribuição de Modelos de Séries Temporais para a Previsão da Arrecadação de ISS.** Dissertação de Mestrado. Instituto de Economia da Universidade Estadual de Campinas, UNICAMP. Campinas, SP. 2003. 123p.

RUSPINI, E.H. ***A New Approach to Clustering.*** Information and Control, vol. 15, n. 1, 1969. p. 22-32.

SANDRI, C.; CORREA, C. **Lógica Nebulosa.** V Escola de Redes Neurais, Promoção: Conselho Nacional de Redes Neurais, ITA. São José dos Campos, SP. 1999. p. c073-c090.

SANTOS, A.A.P. **Previsão Não-Linear da Taxa de Câmbio Real/Dólar Utilizando Redes Neurais e Sistemas Nebulosos.** Dissertação de Mestrado. Programa de Pós-graduação em Economia. Universidade Federal de Santa Catarina, UFSC. Florianópolis, SC. 2005. 98p.

SCIONTI, M.; LANSLOTS, J.P. ***Stabilization Diagrams: Pole Identification using Fuzzy Clustering Techniques.*** Advances in Engineering Software, vol. 36, n. 11-12, 2005. p. 768-779.

SILVA, L.R.S. **Aprendizagem Participativa em Agrupamento Nebuloso de Dados.** Dissertação de Mestrado. Faculdade de Engenharia Elétrica e de Computação. Universidade Estadual de Campinas, UNICAMP. Campinas, SP. 2003. 141p.

SIMÕES, R.F. **Complexos Industriais no Espaço: Uma Análise de Fuzzy Cluster.** Trabalho Técnico. Centro de Desenvolvimento e Planejamento Regional. Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais, UFMG. Belo Horizonte, MG. 2003. 26p.

SONG, Q.; CHISSOM, B.S. ***Forecasting Enrollments with Fuzzy Time Series – Part I.*** Fuzzy Sets Systems, vol. 54, n. 1, 1993. p. 1-10.

SONG, Q.; CHISSOM, B.S. ***Forecasting Enrollments with Fuzzy Time Series – Part II.*** Fuzzy Sets Systems, vol. 62, n. 1, 1994. p. 1-8.

SOUZA, F.J.; FONSECA, O.L H.; MOURA NETO, F.D. ***Credit Analysis Models Using Machine Learning Techniques.*** Proceedings of the 60th International Atlantic Economic Conference, NY, USA. 2005. 6p.

SOUZA, R.R. **Panorama, Oportunidades e Desafios para o Mercado Mundial de Álcool Automotivo.** Dissertação de Mestrado. Programas de Pós-Graduação de Engenharia. Universidade Federal do Rio de Janeiro, UFRJ. Rio de Janeiro, RJ. 2006. 138p.

SUGENO, M. ***Fuzzy Modeling and Control: Selected Works of M. Sugeno.*** CRC Press, NW, USA. 1999. 429p.

TAKAGI, T.; SUGENO, M. ***Fuzzy Identification of Systems and Its Applications to Modeling and Control.*** IEEE Transactions on Systems, Man and Cybernetics, vol. 15, n. 1, 1985. p. 116-132.

TAKATSU, H.; ITOH, T. ***Future Needs for Control Theory in Industry – Reports of the Control Technology Survey in Japanese Industry.*** IEEE Transactions on Control Systems Technology, vol. 7, n. 3. 1999. p. 298-305.

TRABELSI, A.; LAFONT, F.; KAMOUN, M.; ENEA, G. ***Identification of Nonlinear Multivariable Systems by Adaptive Fuzzy Takagi-Sugeno Model.*** International Journal of Computational Cognition, vol. 2, n. 3, 2004. p. 137-158.

VALDES, J. ***Interpreting Fuzzy Clustering Results with Virtual Reality-Based Visual Data Mining: Application to Micro array Gene Expression Data.*** Proceedings of North American Fuzzy Information Processing Society (NAFIPS). Ottawa, Canadá. 2004. 7p.

VALE, M.N. **Agrupamentos de Dados: Avaliação de Métodos e Desenvolvimento de Aplicativo para Análise de Grupos.** Tese de Doutorado. Departamento de Engenharia Elétrica. Pontifícia Universidade Católica do Rio de Janeiro, PUC. Rio de Janeiro, RJ. 2005. 120p.

VAN LITH, P.F.; BETLEM, B.H.L., ROFFEL, B. ***Fuzzy Clustering, Genetic Algorithms and Neuro-Fuzzy Methods Compared for Hybrid Fuzzy-First Principles Modeling.*** Systems Analysis Modelling Simulation, vol. 42, n. 4, 2002. p. 597-630.

VERBEEK, J.J. ***Mixture Models for Clustering and Dimension Reduction.*** Tese de Doutorado. Universiteit van Amsterdan. Amsterdan, Holanda, 2004. 169p.

VERNIEUWE, H.; GEORGIEVA, O.; BAETS, B.; PAUWELS, V.R.N.; VERHOEST, N.E.C.; TROCH, F.P. ***Comparison of Data-Driven Takagi-Sugeno Models of Rainfall-Discharge Dynamics.*** Journal of Hydrology, vol. 302, n. 1-4. 2005. p. 173-186.

VILJAMAA, P. ***Fuzzy Gain Scheduling and Tuning of Multivariable Fuzzy Control – Methods of Fuzzy Computing in Control Systems.*** Tese de Doutorado. Tampere University of Technology. Tampere, Finlândia, 2000. 131p.

WANG, C.H. ***Predicting Tourism Demand using Fuzzy Time Series and Hybrid Grey Theory.*** Tourism Management, vol. 25, n. 3, 2004. p. 367-374.

WANG, H.W.; GU, H.; WANG, Z.L. ***Fuzzy Prediction of Chaotic Time Series Based on SVD Matrix Decomposition.*** Proceedings of the Fourth IEEE International Conference on Machine Learning and Cybernetics. vol. 4. 2005. p. 2493-2498.

XU, Z.; KHOSHGOFTAAR, T.M. ***Identification os Fuzzy Models of Software Cost Estimation***. Fuzzy Sets and Systems, vol.145, n. 1, 2004. p. 141-163.

XU, L.; YAN, Y.; CORNWELL, S.; RILEY, G. ***On-Line Fuel Identification Using Signal Processing and Fuzzy Inference Techniques.*** IEEE Transactions on Instrumentation and Measurement, vol. 53, n. 4, 2004. p. 1316-1320.

YANG, M.S.; YU, N.Y. ***Estimation of Parameters in Latent Class Models using Fuzzy Clustering Algorithms.*** European Journal of Operational Research, vol. 160, n. 2, 2005. p. 515-531.

YE, L.; SPETSAKIS, M.E. ***Clustering on Unobserved Data Using Mixture of Gaussians***. Technical Report. Department of Computer Science. Ontario, Canadá. 2003. 25p.

YEN, J.; LANGARI, R. ***Fuzzy Logic: Intelligence, Control, and Information.*** Prentice Hall, Upper Saddle River, NJ, USA. 1999. 548p.

YU, H.K. ***Weighted Fuzzy Time Series Models for TAIEX Forecasting.*** Physica A, vol. 349, n. 3-4, 2005. p. 609-624.

ZADEH, L.A. ***Fuzzy Sets.*** Information and Control, vol. 8, 1965. p. 338-353.

ANEXO A – REGRAS GERADAS NO CAPÍTULO 6

Tabela A.1: Regras geradas pelo modelo nebuloso TS com algoritmo MGG – Série Álcool Hidratado.

Regra 1:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}799559)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}799559)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 2:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}491362)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}491362)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 3:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}532534)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}532534)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 4:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}686325)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}686325)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 5:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}709502)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-709502)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$

Regra 6:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}798056)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}798056)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 7:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}522694)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}522694)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 8:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}533793)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}533793)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 9:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}640615)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}640615)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 10:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}709379)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}709379)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 11:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}799559)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}799559)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$

Regra 12:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}491362)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}491362)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 13:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}532534)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}532534)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 14:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}686325)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}686325)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 15:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}709502)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-709502)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 16:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}798056)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}798056)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 17:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}522694)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}522694)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$

Regra 18:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}533793)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}533793)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 19:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}640615)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}640615)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 20:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}709379)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}709379)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 21:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}799559)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}799559)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 22:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}491362)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}491362)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 23:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}532534)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}532534)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$

Regra 24:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}686325)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-0{,}686325)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$
Regra 25:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-0{,}709502)}{1}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-709502)}{1}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+44{,}963854)(+315{,}595947)(-459{,}511362)(+415{,}082202)(-312{,}519195)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,5$

$\theta_1 = +44{,}963854; \; \theta_2 = +315{,}595947; \; \theta_3 = -459{,}511362; \; \theta_4 = +415{,}082202; \; \theta_5 = -312{,}519195$

Centros das funções de pertinência do SN:

Centro da função de pertinência Gaussiana 1 da entrada 1 = 0,799559
Centro da função de pertinencia Gaussiana 2 da entrada 1 = 0,522694
Centro da função de pertinência Gaussiana 3 da entrada 1 = 0,532534
Centro da função de pertinência Gaussiana 4 da entrada 1 = 0,640615
Centro da função de pertinência Gaussiana 5 da entrada 1 = 0,709502
Centro da função de pertinência Gaussiana 1 da entrada 2 = 0,798056
Centro da função de pertinência Gaussiana 2 da entrada 2 = 0,491362
Centro da função de pertinência Gaussiana 3 da entrada 2 = 0,533793
Centro da função de pertinência Gaussiana 4 da entrada 2 = 0,686325
Centro da função de pertinência Gaussiana 5 da entrada 2 = 0,709379

Tabela A.2: Regras geradas pelo modelo nebuloso TS com algoritmo GMMEM (Elíptica) – Série da Soja.

Regra 1:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-18{,}870997)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-18{,}870997)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 2:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-14{,}425384)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-14.{,}25384)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 3:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-28{,}069153)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-28{,}069153)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 4:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-18{,}888533)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-18{,}888533)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 5:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-18{,}870997)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-18{,}870997)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 6:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-14{,}425384)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-14{,}425384)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$

Regra 7:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-28{,}069153)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-28{,}069153)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 8:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-18{,}888533)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-18{,}888533)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 9:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-18{,}870997)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-18{,}870997)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 10:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-14{,}425384)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-14{,}425384)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 11:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-28{,}069153)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-28{,}069153)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 12:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-18{,}888533)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-18{,}888533)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$

Regra 13:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-18{,}870997)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-18{,}870997)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 14:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-14{,}425384)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-14{,}425384)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 15:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-28{,}069153)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-28{,}069153)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$
Regra 16:	$SE: y(t-1) = \exp\left[-\frac{1}{2}\frac{(y(t-1)-18{,}888533)}{2500}\right]; \quad y(t-2) = \exp\left[-\frac{1}{2}\frac{(y(t-2)-18{,}888533)}{2500}\right]$ $ENTÃO: \hat{y} = \frac{\left[\sum_j^K A_j(z)\right](+2{,}2128\text{E}+5)(-0{,}0024\text{E}+5)(+0{,}0025\text{E}+5)(-2{,}2122\text{E}+5)}{\sum_j^K A_j(z)}; \; j = 1,\cdots,4$

$\theta_1 = +2{,}2128\text{E}+5; \quad \theta_2 = -0{,}0024+5; \quad \theta_3 = +0{,}0025+5; \quad \theta_4 = -2{,}2122+5;$

Centros das funções de pertinência do SN:

Centro da função de pertinência Gaussiana 1 da entrada 1 = 18,870997
Centro da função de pertinência Gaussiana 2 da entrada 1 = 14,435461
Centro da função de pertinência Gaussiana 3 da entrada 1 = 28,069153
Centro da função de pertinência Gaussiana 4 da entrada 1 = 18,897883
Centro da função de pertinência Gaussiana 1 da entrada 2 = 18,931243
Centro da função de pertinência Gaussiana 2 da entrada 2 = 14,425384
Centro da função de pertinência Gaussiana 3 da entrada 2 = 28,117121
Centro da função de pertinência Gaussiana 4 da entrada 2 = 18,888533