



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO
DEPARTAMENTO DE ENGENHARIA DE COMPUTAÇÃO E
AUTOMAÇÃO INDUSTRIAL

Modelagem Fuzzy Funcional Evolutiva Participativa

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica.
Área de concentração: Automação Industrial

Autor: Elton Mário de Lima

Orientador: Prof. Dr. Fernando Antônio Campos Gomide

Co-Orientadora: Prof^a Dr^a Rosangela Ballini

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

L628m Lima, Elton Mário
 Modelagem fuzzy funcional evolutiva participativa /
 Elton Mário Lima. --Campinas, SP: [s.n.], 2008.

 Orientadores: Fernando Antônio Campos Gomide,
 Rosangela Ballini

 Dissertação (Mestrado) - Universidade Estadual de
Campinas, Faculdade de Engenharia Elétrica e de
Computação.

 1. Sistemas nebulosos. 2. Sistemas inteligentes. 3.
Algoritmos evolutivos. 4. Previsão de séries temporais. I.
Gomide, Fernando Antônio Campos. II. Ballini, Rosangela.
III. Universidade Estadual de Campinas. Faculdade de
Engenharia Elétrica e de Computação. IV. Título.

Título em Inglês: Evolving participatory learning fuzzy modeling

Palavras-chave em Inglês: Fuzzy systems, Fuzzy clustering, Fuzzy dynamic modeling,
real-time systems, Time series prediction, Evolving
intelligent systems

Área de concentração: Automação

Titulação: Mestre em Engenharia Elétrica

Banca examinadora: Benjamin Menezes, Takaaki Ohishi

Data da defesa: 04/07/2008

Programa de Pós-Graduação: Engenharia Elétrica

Elton Mário de Lima

Modelagem Fuzzy Funcional Evolutiva Participativa

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica.
Área de concentração: Automação Industrial
Aprovação em 04/07/2008

Banca Examinadora:

Prof. Dr. Benjamin Menezes – UFMG

Prof. Dr. Fernando Antônio Campos Gomide – UNICAMP

Prof. Dr. Takaaki Ohishi – UNICAMP

Universidade Estadual de Campinas
Campinas, SP

COMISSÃO JULGADORA - TESE DE MESTRADO

Candidato: Elton Mário de Lima

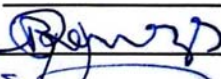
Data da Defesa: 4 de julho de 2008

Título da Tese: "Modelagem Fuzzy Funcional Evolutiva Participativa"

Prof. Dr. Fernando Antônio Campos Gomide (Presidente):



Prof. Dr. Benjamim Rodrigues de Menezes:



Prof. Dr. Takaaki Ohishi:



Elton Mário de Lima

Evolving Participatory Learning Fuzzy Modeling

Master of Science Thesis
State University of Campinas
Faculty of Electrical and Computer Engineering

Campinas, SP
Brazil
2008

Aos meus pais Gilson e Nilze Lima.

Agradecimentos

Aos meus pais Gilson e Nilze pela educação, irrestrito apoio e confiança.

Aos professores Fernando Gomide e Rosangela Ballini pelos ensinamentos, tolerância e exemplar orientação.

Aos colegas de formatura e pós-graduação pelos momentos divertidos.

Aos colegas do Cose pela ajuda e companhia acadêmica.

À UNICAMP e Fapesp por toda estrutura e apoio financeiro.

Aos professores Benjamin Menezes, Takaaki Ohishi e Peter Sussner pelas valiosas sugestões.

Resumo

Este trabalho propõe um modelo fuzzy funcional evolutivo que utiliza uma aplicação do aprendizado participativo para a construção de uma base de regras. O aprendizado participativo é um modelo de aprendizado baseado na noção de compatibilidade para a atualização do conhecimento do sistema. O aprendizado participativo pode ser traduzido em um algoritmo de agrupamento não supervisionado conhecido como agrupamento participativo. O algoritmo intitulado Aprendizado Participativo Evolutivo é proposto para construir um modelo fuzzy funcional evolutivo no qual as regras são obtidas a partir de um algoritmo de agrupamento não supervisionado. O algoritmo utiliza uma versão do agrupamento participativo para a determinação de uma base de regras correspondente ao modelo funcional do tipo Takagi-Sugeno evolutivo. A partir de uma noção generalizada, o modelo proposto é aplicado em problemas de previsão de séries temporais e os resultados são obtidos para a conhecida série Box-Jenkins, além da previsão de uma série de carga horária de energia elétrica. Os resultados são comparados com o modelo Takagi-Sugeno evolutivo que utiliza a noção de função potencial para agrupar os dados dinamicamente e com duas abordagens baseadas em redes neurais. Os resultados mostram que o modelo proposto é eficiente e parcimonioso, abrindo potencial para aplicações e estudos futuros.

Palavras-chave: Sistemas nebulosos, Sistemas inteligentes, Algoritmos evolutivos, Previsão de séries temporais.

Abstract

This work introduces an approach to develop evolving fuzzy rule-based models using participatory learning. Participatory learning assumes that learning and beliefs about a system depend on what the learning mechanism knows about the system itself. Participatory learning naturally augments clustering and yields an effective unsupervised fuzzy clustering algorithms for on-line, real time domains and applications. Clustering is an essential step to construct evolving fuzzy models and plays a key role in modeling performance and model quality. A least squares recursive approach to estimate the consequent parameters of the fuzzy rules for on-line modeling is emphasized. Experiments with the classic Box-Jenkins benchmark are conducted to compare the performance of the evolving participatory learning with the evolving fuzzy system modeling approach and alternative fuzzy modeling and neural methods. The experiments show the efficiency of evolving participatory learning to handle the benchmark problem. The evolving participatory learning method is also used to forecast the average hourly load of an electric generation plant and compared against the evolving fuzzy system modeling using actual data. The results confirm the potential of the evolving fuzzy participatory method to solve real world modeling problems.

Keywords: Fuzzy systems, Fuzzy clustering, Fuzzy dynamic modeling, Real-time systems, Time series prediction.

Conteúdo

Lista de Figuras	17
Lista de Símbolos	19
Trabalhos Publicados	21
1 Introdução	23
1.1 Motivação	23
1.2 Revisão da Literatura	24
1.3 Objetivos	27
1.4 Organização do Trabalho	28
2 Aprendizado Participativo	31
2.1 Introdução	31
2.2 Formulação do Aprendizado Participativo	32
2.3 Análise do Aprendizado Participativo	36
2.4 Resumo	39
3 Convergência do Aprendizado Participativo	41
3.1 Introdução	41
3.2 Caso Particular	42
3.3 Caso Geral	47
3.4 Resumo	51
4 Agrupamento Participativo	53
4.1 Introdução	53
4.2 Algoritmo de Agrupamento Participativo	54
4.3 Parâmetros do Agrupamento Participativo	59
4.4 Resumo	63

5	Modelos Fuzzy Funcionais Evolutivos	65
5.1	Introdução	65
5.2	Modelos Fuzzy Funcionais Takagi-Sugeno	67
5.3	Modelo Takagi-Sugeno Evolutivo	68
5.4	Resumo	73
6	Modelo Takagi-Sugeno Evolutivo Participativo	75
6.1	Introdução	75
6.2	Aprendizado Participativo Evolutivo	76
6.2.1	Atualização dos antecedentes das regras	77
6.2.2	Atualização dos conseqüentes das regras	81
6.3	Algoritmo de Aprendizado Participativo Evolutivo	85
6.4	Resumo	90
7	Resultados	91
7.1	Introdução	91
7.2	Previsão da Série Box-Jenkins	93
7.3	Previsão de Carga Elétrica	96
7.4	Resumo	102
8	Conclusão	103
	Referências Bibliográficas	105

Lista de Figuras

2.1	Mecanismo do aprendizado participativo.	35
2.2	Comportamento do índice de alerta em relação ao parâmetro β	37
2.3	Comportamento do termo $G^k = (\rho^k)^{1-\alpha^k}$	38
3.1	Ilustração da convergência da sequência $\{v^k\}$	45
3.2	Exemplo com $f(v) = \text{sen}(v) + v$ e $v^{k+1} = v^k + \alpha G^k \text{seno}(v^k)$	46
4.1	Índice de compatibilidade entre centros.	56
4.2	Redefinição dos grupos redundantes.	57
4.3	Caso onde um novo grupo não será redundante.	60
4.4	Caso onde é melhor atualizar um grupo existente.	61
5.1	Espaço dos antecedentes e conseqüentes.	66
5.2	Mecanismo da base de regras no eTS.	69
5.3	Ilustração da idéia de função potencial.	70
6.1	Mecanismo da base de regras no APE.	76
6.2	Função de pertinência Gaussiana com valor modal no centro de grupo. . .	81
6.3	Diagrama com os passos do APE.	88
7.1	Escolha do parâmetro λ	93
7.2	Comportamento do índice de alerta na previsão da série Box-Jenkins. . .	95
7.3	Saída do APE na previsão da série Box-Jenkins.	95
7.4	Função de autocorrelação parcial da série horária de carga elétrica. . . .	97
7.5	Comportamento do índice de alerta na série horária de carga elétrica. . . .	98
7.6	Saída do APE na previsão da série horária de carga elétrica.	99
7.7	Saída do eTS na previsão da série horária de carga elétrica.	99
7.8	Saída do xTS na previsão da série horária de carga elétrica.	100

7.9 Saída da MLP na previsão da série horária de carga elétrica.	100
7.10 Saída do ANFIS na previsão da série horária de carga elétrica.	100
7.11 Saídas para diversos períodos de treinamento	101

Lista de Símbolos

$[0, 1]$	- Intervalo unitário real
$[0, 1]^p$	- Hipercubo unitário real p -dimensional
k	- Instante, estágio ou etapa no qual o dado x^k é fornecido
x^k	- Vetor de dados em k , p -dimensional
$\mathcal{R}_i^k$	- Regra i em k
c^k	- Número de grupos ou regras em k
v_i^k	- Centro do grupo i em k
ρ_i^k	- Índice de compatibilidade entre a observação x^k e o centro v_i^k
ρ_{ij}^k	- Índice de compatibilidade entre os centros v_i^k e v_j^k
a_i^k	- Índice de alerta do grupo i em k
β	- Sensibilidade do índice de alerta
α	- Taxa de aprendizagem
τ	- Parâmetro limiar para o índice de alerta
λ	- Parâmetro limiar para o índice de compatibilidade entre os centros dos grupos
A_i^k	- Vetor de funções de pertinência do antecedente da regra i em k
γ_i^k	- Vetor de parâmetros do conseqüente da regra i em k
μ_i^k	- Grau de ativação da regra i em k
y_i^k	- Saída da regra i em k
y^k	- Saída da base de regras
P^k	- Função potencial calculada no ponto x^k
$P(v_i^k)$	- Função potencial calculada no centro v_i^k

Trabalhos Publicados

Os artigos elaborados e apresentados durante o período de desenvolvimento do presente trabalho foram os seguintes:

Internacionais

1. Lima, E., Gomide, F. e Ballini, R. (2006). Participatory evolving fuzzy modeling, *Proceedings of the International Symposium on Evolving Fuzzy Systems – EFS’06*, IEEE Press, Ambleside, Lake District, UK, pp. 36–41.
- A apresentação oral desse trabalho originou o prêmio de *Best Student Paper*.
2. Lima, E., Hell, M., Ballini, R. e Gomide, F. (2008). *Evolving Fuzzy Modeling Using Participatory Learning*, em *Evolving Intelligent Systems: Methodology and Applications*, Angelov, P., Filev, D. e Kasabov, N. (Eds) , Wiley-Interscience/IEEE Press.

Nacionais

1. Lima, E., Ballini, R. e Gomide, F. (2006). Modelagem evolucionária participativa nebulosa, *XVI Congresso Brasileiro de Automática – CBA’06*, Salvador, BA, Brasil, pp. 2519–2524.
2. Lima, E., Ballini, R. e Gomide, F. (2007). Modelagem de Sistemas Utilizando Aprendizado Participativo, *VIII Congresso Brasileiro de Redes Neurais – CBRN’07*, Florianópolis, SC, Brasil, pp. 2619–2624.

Capítulo 1

Introdução

1.1 Motivação

A busca pelo desenvolvimento de soluções de problemas não lineares e de difícil compreensão é um desafio presente em diversas áreas do conhecimento humano. Seja em questões de engenharia, física, economia, e mesmo biologia, desenvolver um modelo capaz de descrever sistemas dinâmicos não lineares representa uma tarefa desafiadora. A presença de incertezas e propriedades que não podem ser bem estabelecidas são exemplos de dificuldades encontradas quando busca-se elaborar um mecanismo para a modelagem de sistemas. Espera-se que o modelo seja capaz de adaptar-se à diferentes situações (Nelles 2000), acumulando inteligência.

Partindo da definição presente em (Michaelis 1998), a palavra evolutivo tem, entre outros, os seguintes significados:

- i) progresso paulatino e contínuo a partir de um estado inferior ou simples para um superior, mais complexo ou melhor;
- ii) transformação lenta, em leves mudanças sucessivas.

Os algoritmos evolutivos podem ser definidos como algoritmos cujos parâmetros são modificados gradativamente, adaptando a apresentação de determinado fenômeno (Angelov e Kasabov 2006). Redes neurais artificiais e modelos fuzzy (Pedrycz e Gomide 2007) são exemplos de modelos que podem ser desenvolvidos evolutivamente.

Modelos fuzzy, por sua vez, podem ser caracterizados como modelos lingüísticos ou modelos funcionais. Modelos fuzzy lingüísticos são modelos baseados em regras fuzzy onde tanto o antecedente como o conseqüente são definidos por termos lingüísticos. Nos modelos fuzzy funcionais, o antecedente é definido por termos lingüísticos, mas o conseqüente é uma função das variáveis do antecedente (Pedrycz e Gomide 2007).

Neste trabalho, serão considerados os Modelos Fuzzy Funcionais Evolutivos. Modelos fuzzy funcionais evolutivos utilizam um mecanismo onde a base de regras pode ser expandida ou reduzida, alterando-se o número de regras que descrevem o modelo. A inteligência é desenvolvida à cada novo dado de entrada, criando uma nova regra, atualizando algumas das regras existentes e excluindo regras redundantes.

Nos modelos fuzzy funcionais conhecidos como Takagi-Sugeno, os conseqüentes são funções que dependem da variável de entrada, e a saída da base de regras é obtida a partir dessas funções e dos respectivos graus de ativação (Takagi e Sugeno 1985), (Nguyen e Sugeno 1998). Os antecedentes podem ser determinados a partir de um algoritmo de agrupamento.

Modelos fuzzy funcionais evolutivos vêm sendo aplicados nos últimos anos na solução de diversos problemas, como identificação, controle, navegação autônoma e previsão de séries temporais (Angelov, Filev e Kasabov (Eds) 2008). Os algoritmos associados à modelos fuzzy funcionais evolutivos definem uma promissora abordagem para a construção de modelos não lineares e adaptativos.

1.2 Revisão da Literatura

A literatura recente destaca diversas propostas de algoritmos evolutivos, capazes de interpretar gradativamente um conjunto de dados e definir a decisão a ser tomada à cada etapa (Yen, Wang e Gillespie 1998), (Kasabov 2001), (Angelov e Filev 2004), (Kasabov e Filev 2006), (Angelov e Kasabov 2006).

Durante os anos 90, novas abordagens começaram a utilizar a extração de padrões em conjuntos de dados, aplicando ajustes paramétricos locais (Geva 1999). Destacam-se ainda o uso de algoritmos de agrupamento (Bezdek 1981), (Kaymak e Setnes 2000), mínimos quadrados (Young 1984) e otimização não linear (Bazaraa, Sherali e Shetty 2006). Essas abordagens têm como objetivo construir um sistema capaz de identificar um conjunto de dados e compor um modelo que descreva esse conjunto de forma eficiente (Chiu 1994), (Nguyen e Sugeno 1998).

Muitos dos trabalhos publicados na área de modelos adaptativos utilizam conjuntos nebulosos (Zadeh 1965), (Zadeh 1973) redes neurais artificiais (Werbos 1990), (Haykin 1998), e ainda algoritmos genéticos (Holland 1975), (Goldberg 1989), (Koza 1992). Recentemente diversos avanços foram divulgados, destacando o uso de redes neurais fuzzy (Lin, Lin e Shen 2001), (Kasabov e Song 2002) e modelos que utilizam regras fuzzy funcionais (Angelov, Xydeas e Filev 2004), (Angelov e Filev 2005).

O sistema de inferência neurofuzzy (ANFIS), proposto em (Jang 1993), é um modelo neural artificial fuzzy no qual o ajuste dos parâmetros é baseado no algoritmo do gradiente descendente. Este modelo tem sido amplamente utilizado em diversas aplicações como, por exemplo, modelagem de sistemas não lineares (Buragohain e Mahanta 2008) e previsão de séries temporais (Chen e Zhang 2005). Além do ANFIS, destacam-se modelos neurais fuzzy aplicados na identificação e controle de sistemas (Lee, Lee e Park 1995), (Figueiredo e Gomide 1998), previsão de séries temporais (Figueiredo, Ballini, Soares, Andrade e Gomide 2004) e também trabalhos utilizando memórias associativas e morfologia (Sussner e Valle 2007) e classificação de padrões (Brouwer 2000), (Sussner e Valle 2007).

Devido à complexidade dos problemas reais, os quais são em geral não lineares, saber com antecedência o número adequado de regras fuzzy que irão compor o modelo torna-se inviável. Normalmente, este problema é tratado por tentativa e erro, sendo que estruturas de diferentes tamanhos são construídas, ajustadas e avaliadas. O modelo com melhor desempenho é escolhido como o mais adequado.

Como alternativa às abordagens neurofuzzy surgem os algoritmos com aprendizado seqüencial, conhecidos na literatura como modelos de aprendizado evolutivo. Nesta abordagem a atualização é feita em cada instante a partir dos novos dados que são apresentados ao modelo. A estrutura é modificada acrescentando ou eliminando regras fuzzy, de acordo com critérios previamente estabelecidos. Em outras palavras, tanto a estrutura quanto os parâmetros não são fixos, podendo variar com o tempo. Os trabalhos desenvolvidos em (Kasabov 2001), (Kasabov e Song 2002), (Er e Wu 2002), (Angelov e Filev 2004), (Angelov e Filev 2005), (Leng, McGinnity e Prasad 2005) e (Lima, Gomide e Ballini 2006) são alguns exemplos deste tipo de algoritmos adaptativos.

No trabalho desenvolvido por N. Kasabov (2001) é proposto um modelo fuzzy neural evolutivo, o qual tem como característica principal a capacidade de aprendizado, assim como o ajuste local dos elementos que compõem a estrutura. A estrutura do modelo é modificada a cada instante de tempo utilizando um algoritmo de aprendizado incremental. Uma extensão deste trabalho é descrita em (Kasabov e Song 2002), onde a saída estimada é calculada considerando apenas as regras com maior grau de ativação para cada dado de entrada apresentado.

O modelo proposto por M. Er e S. Wu (2002) sugere um algoritmo evolutivo, onde a estrutura é modificada por meio da inserção de novas regras fuzzy quando necessário. Após o período de aprendizado evolutivo, o processo é interrompido e a relevância de cada uma das regras é verificada utilizando critérios de parada. Da mesma forma, o modelo Takagi-Sugeno evolutivo (eTS), proposto por P. Angelov e D. Filev (2004) considera apenas a geração de novas regras durante o aprendizado, o que em alguns casos leva à obtenção de estruturas complexas.

Por outro lado, em propostas seguintes como as descritas em (Angelov e Filev 2005), (Leng et al. 2005) e (Lima, Ballini e Gomide 2006) os modelos são modificados durante o aprendizado evolutivo, de tal forma que tanto os critérios de adição como de redução de regras são verificados constantemente, fornecendo assim estruturas mais

compactas e reduzindo o número de parâmetros a serem ajustados a cada instante.

Dessa forma, problemas de alta complexidade e com solução analítica desconhecida podem ser analisados sob um novo ponto de vista, determinando um grande avanço em diversas das áreas já citadas, dentre elas previsão de séries temporais (Filev e Angelov 1992), (Geva 1999), (Kasabov e Song 2002), (Angelov e Filev 2004), (Sussner e Valle 2007), (Lima, Ballini e Gomide 2007).

1.3 Objetivos

O objetivo deste trabalho é apresentar um algoritmo que utiliza o conceito de Aprendizagem Participativa (Yager 1990) para a construção de um Modelo Fuzzy Funcional Evolutivo (Angelov e Filev 2004). Em modelos fuzzy funcionais evolutivos, as regras podem ser obtidas a partir de um algoritmo de agrupamento não supervisionado. Com isso, o algoritmo proposto intitulado Aprendizado Participativo Evolutivo utiliza uma versão do agrupamento participativo (Silva 2003) para a determinação de uma base de regras Takagi-Sugeno evolutiva (Yen et al. 1998).

Com uma abordagem inspirada no modelo Takagi-Sugeno evolutivo conhecido como eTS (Angelov 2002), o Aprendizado Participativo Evolutivo (APE) proposto neste trabalho utiliza um algoritmo de agrupamento não supervisionado para a construção dos antecedentes de uma base de regras evolutiva. Definido os antecedentes, um modelo linear é ajustado para cada grupo utilizando o algoritmo de mínimos quadrados recursivos (Young 1984).

Assim, o objetivo geral deste trabalho consiste no desenvolvimento de um algoritmo de aprendizado participativo evolutivo, tendo como principais características:

- Capacidade de extração de conhecimento;

- Capacidade de adaptação do modelo;
- Estruturas parcimoniosas;
- Algoritmo de aprendizado computacionalmente eficiente.

Como objetivo específico deste trabalho, considera-se a aplicação do algoritmo proposto na previsão de um conjunto de séries temporais. As aplicações que serão objeto de estudo são a previsão da série Box-Jenkins e previsão da série de carga horária de energia elétrica.

O desempenho do algoritmo proposto neste trabalho é comparado com outras abordagens presentes na literatura, como o algoritmo eTS proposto por (Angelov e Filev 2004), e modelos baseados em redes neurais.

1.4 Organização do Trabalho

Após esta introdução, o trabalho discute o conceito do aprendizado participativo (Yager 1990) no Capítulo 2. Considerações acerca desse paradigma e da respectiva representação matemática são feitas no Capítulo 3, incluindo uma análise sobre a convergência utilizando resultados conhecidos na teoria do ponto fixo (Lages 2003).

Em seguida, o Capítulo 4 descreve o agrupamento participativo, resultado da aplicação da noção de aprendizagem participativa em problemas de agrupamento de dados. Nesse capítulo, uma versão do algoritmo do agrupamento participativo é definida, e uma análise da influência dos respectivos parâmetros do algoritmo também é feita.

No Capítulo 5 são revistos os modelos fuzzy funcionais. O modelo Takagi-Sugeno evolutivo (eTS), que utiliza um algoritmo de agrupamento baseado na noção de função potencial para a atualização dos centros, é detalhado juntamente com as idéias principais propostas em (Angelov e Filev 2004).

O Capítulo 6 trata da contribuição principal deste trabalho, o algoritmo de Aprendizado Participativo Evolutivo. Para tal, o agrupamento participativo e os conceitos de regras funcionais definidos nos capítulos anteriores são utilizados.

O Aprendizado Participativo Evolutivo é aplicado em problemas de previsão de séries temporais, cujos resultados são apresentados no Capítulo 7. Comparações com outros modelos presentes na literatura como o modelo eTS (Angelov 2002) e duas outras variações conhecidas como simpl_eTS (Angelov e Filev 2005) e xTS (Angelov e Zhou 2006) são apresentadas. Resultados obtidos a partir de uma rede neural multicamada MPL (Rumelhart, Hinton e Williams 1986) e do sistema de inferência neurouuzzy ANFIS (Jang 1993) também são utilizados como comparação para o desempenho do algoritmo proposto.

O Capítulo 8 conclui o trabalho resumindo suas principais contribuições e propondo os tópicos a serem desenvolvidos futuramente.

Capítulo 2

Aprendizado Participativo

2.1 Introdução

O desenvolvimento de algoritmos capazes de atuar como um mecanismo de aprendizagem despertou bastante interesse nos últimos anos. Além de auxiliar na compreensão de alguns aspectos da aprendizagem humana, a criação de algoritmos que simulem artificialmente um processo de aprendizado é de grande importância, com aplicações em problemas cuja solução analítica não é conhecida.

O entendimento e compreensão do aprendizado humano serve de inspiração para diversos estudos recentes (Kasabov 2001), (Yager 2004). Considerando a estrutura atual como o conhecimento do sistema, cada entrada é vista como uma informação a ser utilizada para atualizar este conhecimento. Dessa forma, busca-se uma maneira de computar e processar informações desenvolvendo gradativamente uma solução para problemas de difícil abordagem analítica.

Nesse contexto, o Aprendizado Participativo (Yager 1990) representa um paradigma importante com diversas aplicações, entre elas a construção de redes neurais (Yager 2004) e algoritmos de agrupamento (Silva 2003), (Silva, Gomide e Yager 2005).

O Aprendizado Participativo (AP) é um mecanismo de aprendizagem no qual a percepção acerca de um sistema depende de quanto o mecanismo é capaz de representar o conhecimento já estabelecido sobre o sistema. Em linhas gerais, o conhecimento atual é parte de um processo de aprendizagem que depende da coerência entre o estado atual do

sistema e a observação mais recente. A renovação do conhecimento, isto é, o aprendizado, é caracterizado pela interpretação da nova informação no contexto atual.

Como idéia central do Aprendizado Participativo, deve-se estabelecer uma conexão entre as informações coletadas a cada instante e o estado atual, sendo que as observações coerentes com o sistema devem atualizá-lo de alguma forma, e observações pouco coerentes precisam ser detectadas e subestimadas no processo afim de rever o que o sistema aprendeu até então, ou proteger o sistema de possíveis erros e divergências. O AP define um processo de aprendizagem passo a passo que pode ser implementado recursivamente (Yager 2004).

2.2 Formulação do Aprendizado Participativo

O aprendizado participativo utiliza a noção de que, quanto maior a compatibilidade entre uma nova observação e o estado atual do sistema, maior será o impacto dessa informação na aprendizagem do mesmo. Nesse sentido, se $\bar{v} \in [0, 1]^p$ é a representação do conhecimento, pode-se dizer que o objetivo é aprender o valor de $\bar{v}$ a partir de um conjunto de observações $x^k \in [0, 1]^p$. Essa aprendizagem é feita via atualização de estimativas v^k feitas de acordo com sua compatibilidade em relação à cada x^k observado.

Em outras palavras, busca-se definir uma sequência $v^k = v(G^k, x^{k-1}, v^{k-1})$ função que, sob certas condições, possua um comportamento convergente e consistente com o sistema objeto de aprendizagem. Para tal, deve-se considerar as informações, a experiência (ou percepção) acumulada, e uma função de compatibilidade que sirva para medir o quanto o conhecimento está de acordo com as observações. A função de compatibilidade pode ser definida como uma função $\rho : [0, 1]^p \times [0, 1]^p \rightarrow [0, 1]$ tal que seja (Klir e Yuan 1995):

- 1) reflexiva: $\rho(x, x) = 1$;
- 2) simétrica: $\rho(x, v) = \rho(v, x)$.

No contexto deste trabalho, $\rho(x, v) = 1$ define compatibilidade plena entre x e v , e $\rho(x, v) = 0$ define incompatibilidade entre x e v .

A função de compatibilidade ou, simplesmente, compatibilidade, representa o quanto o dado observado está de acordo com o modelo atual. Se $\rho \approx 1$, considera-se que o sistema está mais propenso a aprender, ocorrendo o oposto para $\rho \approx 0$. Um mecanismo que implementa essa idéia é o seguinte.

Considere x^k uma variável real em $[0, 1]^p$ e $\alpha \in [0, 1]$ um parâmetro chamado de taxa de aprendizagem. Então a dinâmica associada ao aprendizado participativo é definida por uma sequência $\{v^k\}$ na forma (Yager 1990):

$$v^{k+1} = v^k + \alpha G^k(x^k - v^k) \quad (2.1)$$

no qual $v^1 \in [0, 1]^p$ e $G^k = G(\rho(x^k, v^k))$ é uma função que depende da compatibilidade entre os vetores x^k e v^k em $k = 1, 2, \dots$.

Nesse processo, o conceito participativo é caracterizado pela contribuição de cada observação de acordo com um critério de aceitação, obtido afim de determinar a coerência entre o conhecimento atual expresso por v^k e um novo dado x^k no instante k . Em outras palavras, busca-se definir o quanto a última observação pode contribuir na atualização do conhecimento representado pela variável v^k .

Para o contexto em questão, espera-se que a compatibilidade seja tão maior quanto mais próximo x^k for de v^k . Dentre as diversas escolhas possíveis, uma sugestão é escolher $G^k = \rho^k$, gerando a seguinte expressão (Yager 1990):

$$v^{k+1} = v^k + \alpha \rho^k(x^k - v^k) \quad (2.2)$$

sendo $\alpha \in [0, 1]$ a taxa de aprendizagem e $\rho^k \in [0, 1]$ uma função de compatibilidade.

Há uma infinidade de maneiras para uma construção que satisfaça a definição

de função de compatibilidade. Em particular, utilizar a distância como uma medida da proximidade entre x^k e v^k é uma escolha intuitiva para decidir se as mesmas são ou não compatíveis. Uma escolha simples e direta é:

$$\rho^k = 1 - d(x^k, v^k) \quad (2.3)$$

na qual $d(x^k, v^k)$, ou simplesmente d^k , é a distância entre x^k e v^k . Vale notar que essa abordagem permite o cálculo de diferentes medidas de compatibilidade para cada componente de x^k .

No entanto, tal abordagem não leva em conta situações onde x^k represente uma informação conflitante com v^k , isto é, casos onde observações com baixo índice de compatibilidade são recebidas durante instantes consecutivos. Com isso, é importante que o mecanismo possa detectar também inconsistências no aprendizado, minimizando a influência do índice de compatibilidade na atualização de v^k em casos onde o índice de alerta é alto. Esse índice de alerta é definido de forma que, quanto maior for o seu valor, menor deve ser a influência do índice de compatibilidade na atualização do sistema.

De fato, um baixo valor para o índice de compatibilidade em um determinado instante pode ser visto como uma aversão ao aprendizado nesse instante. Por outro lado, baixos valores de compatibilidade durante vários instantes consecutivos podem representar uma necessidade de revisão no aprendizado, e não necessariamente uma aversão (Yager 2004). Nesse sentido, o índice de alerta deve detectar se uma observação pouco compatível será subestimada, ou se a baixa compatibilidade caracteriza a necessidade de revisão do sistema.

Dentre outras possibilidades, considere o índice de alerta dado por:

$$a^{k+1} = a^k + \beta(1 - \rho^k - a^k) \quad (2.4)$$

sendo $\beta \in [0, 1]$ um parâmetro e $a^k \in [0, 1]$ com $k = 1, 2, \dots$

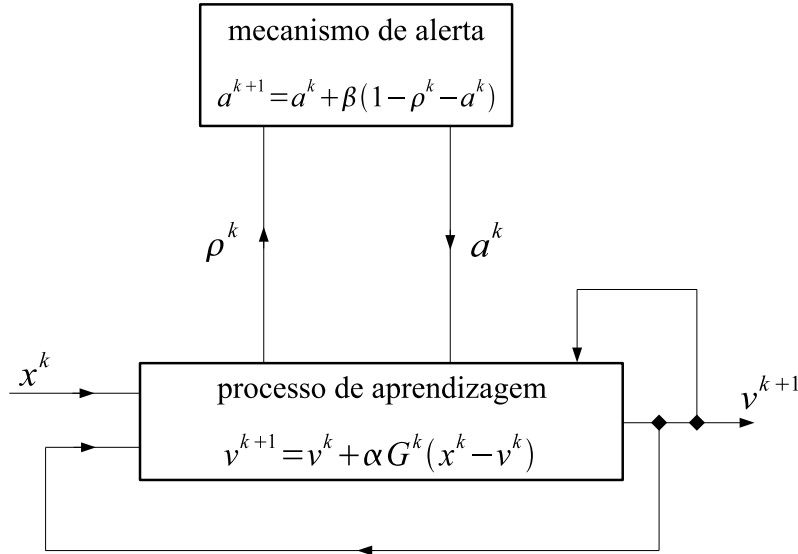


Figura 2.1: Mecanismo do aprendizado participativo.

Na Figura 2.1 é ilustrado que para cada entrada x^k o índice de compatibilidade ρ^k é obtido e enviado ao mecanismo de alerta, e esse mecanismo de alerta monitora o desempenho do aprendizado. O índice de alerta a^k é calculado e o aprendizado é atualizado na direção do vetor $(x^k - v^k)$, onde o tamanho do passo depende do termo αG^k . Quanto maior for o índice de alerta, menor será a influência do índice de compatibilidade no cálculo de v^{k+1} . Ainda pela Figura 2.1, o mecanismo é considerado participativo pois a atualização do sistema é feita de acordo com o conhecimento acumulado, ou seja, a variação no aprendizado depende do quanto já é conhecido acerca do sistema.

A seção seguinte relata como o termo G^k é calculado de acordo com os índices de compatibilidade e alerta ρ^k e a^k , e como os parâmetros $\beta \in [0, 1]$ e $\alpha \in [0, 1]$ são utilizados no aprendizado participativo.

2.3 Análise do Aprendizado Participativo

Esta seção faz uma descrição do comportamento do aprendizado participativo em relação aos parâmetros que o definem. A análise da influência do parâmetro $\alpha \in [0, 1]$ na convergência do aprendizado participativo é feita no Capítulo 3.

Considerando a expressão (2.4), o parâmetro $\beta \in [0, 1]$ controla a taxa de variação do índice de alerta, ou seja, a sensibilidade do mecanismo em relação às variações no dado x^k . Tomando também o índice de compatibilidade (2.3), tem-se a seguinte desigualdade:

$$a^{k+1} - a^k = \beta(d^k - a^k) \quad \Rightarrow \quad a^{k+1} - a^k \leq \beta d^k \quad (2.5)$$

sendo $d^k = d(x^k, v^k)$ a distância entre v^k e a entrada x^k .

Para valores do parâmetro β próximos de 1, a sensibilidade do índice de alerta em relação às variações da entrada na expressão (2.4) é maior. Por outro lado, a desigualdade definida em (2.5) mostra que, para valores de β próximos de 0, as variações no valor do índice de alerta a^k serão menores.

Como ilustrado na Figura 2.2, que mostra o comportamento do índice de alerta a^{k+1} em relação ao parâmetro β , se $\beta = 1$ na expressão (2.4) então o índice de alerta seguinte a^{k+1} será igual à distância d^k entre v^k e a entrada x^k , e se $\beta = 0$ o índice de alerta será constante para todo $k = 1, 2, \dots$

Em outras palavras, (Yager 2004) sugere o parâmetro β como um grau de conservadorismo do sistema, e o índice de alerta a^k é a confiança do sistema em relação ao aprendizado representado por v^k . Assim, para o cálculo do alerta a^{k+1} leva-se em conta tanto a compatibilidade ρ^k no instante atual, quanto o valor anterior a^k .

Enquanto o índice de compatibilidade (2.3) mede a aceitação de uma obser-

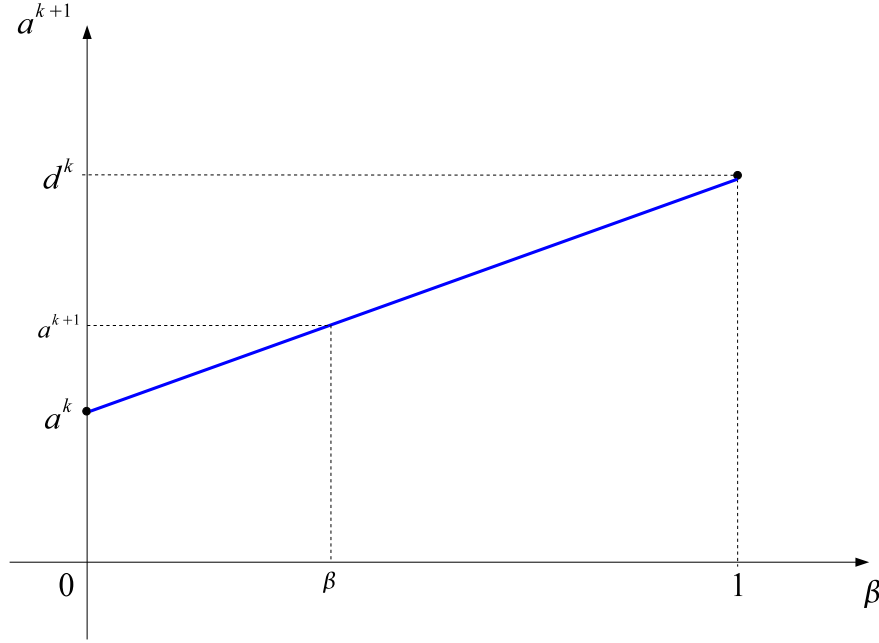


Figura 2.2: Comportamento do índice de alerta em relação ao parâmetro β .

vação, o índice de alerta (2.4) age como um crítico, gerando um procedimento que busca considerar a coerência da nova observação de forma a atualizar o estado atual de acordo com o mecanismo participativo (Yager 1990).

Com isso, a atualização v^{k+1} do aprendizado pode ser ampliada considerando também o índice de alerta. A idéia é que a^k possa influenciar a aprendizagem alterando a influência do índice de compatibilidade ρ^k na expressão definida em (2.1). Dentre as diversas possibilidades, (Yager 2004) sugere $G^k = (\rho^k)^{1-a^k}$, de forma que:

$$v^{k+1} = v^k + \alpha(\rho^k)^{1-a^k}(x^k - v^k) \quad (2.6)$$

O índice de alerta controla o efeito do índice de compatibilidade na aprendizagem. Como $\rho^k \in [0, 1]$ e $a^k \in [0, 1]$, tem-se que $(\rho^k)^{1-a^k} \geq \rho^k$ e a atualização de v^k dada por (2.6) torna-se mais sensível que a definida sem o índice de alerta em (2.2), pois em (2.6) a variação no valor v^k será maior.

A Figura 2.3 ilustra o comportamento do termo $G^k = (\rho^k)^{1-a^k}$. Em particular, para $a^k = 0$ tem-se $(\rho^k)^{1-a^k} = \rho^k$, e a atualização considerando o índice de alerta em (2.6) é idêntica à expressão (2.2) que não considera o índice. Para $a^k = 1$, tem-se $(\rho^k)^{1-a^k} = 1$, ou seja, quando o índice de alerta do modelo é alto, o índice de compatibilidade é suprimido na atualização do aprendizado, pois $v^{k+1} = v^k + \alpha(x^k - v^k)$.

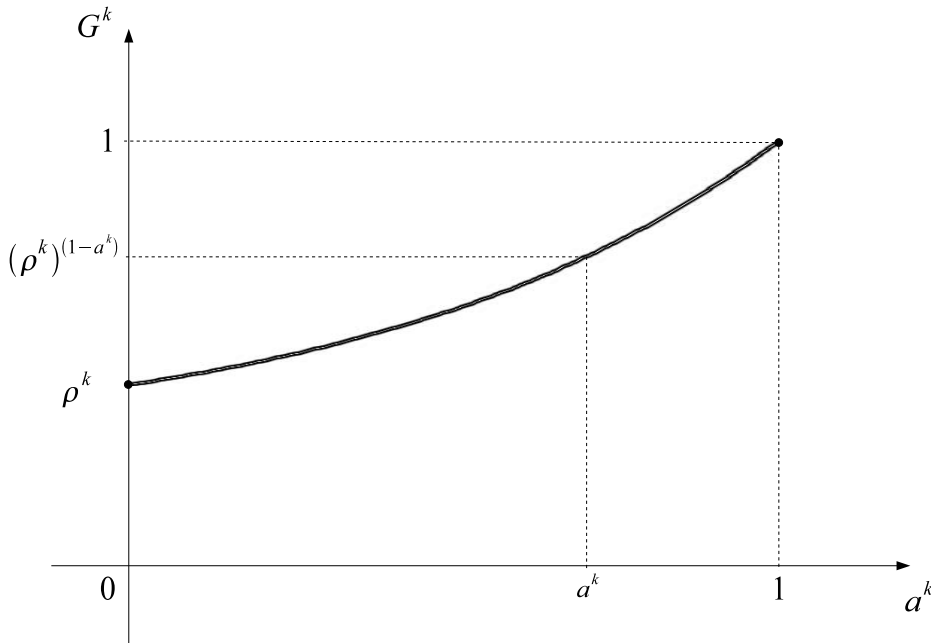


Figura 2.3: Comportamento do termo $G^k = (\rho^k)^{1-a^k}$.

Portanto, um modelo de aprendizado descrito dessa forma representa uma dinâmica iterativa cujo comportamento depende da característica das observações x^k . Se v^k e x^k forem valores determinísticos, então o aprendizado participativo pode ser visto como uma busca iterativa (Conte e de Boor 1981). Caso v^k e x^k sejam variáveis aleatórias, procedimentos de aproximação estocástica auxiliam na análise do aprendizado participativo (Wasan 1969), (Kushner e Clark 1978). Algumas condições necessárias para a convergência de v^k são apresentadas no Capítulo 3 seguinte.

2.4 Resumo

Esse capítulo introduziu o conceito do aprendizado participativo. Proposto em (Yager 1990), esse conceito atualiza o aprendizado de acordo com o conhecimento adquirido até então, a partir de novas observações. As noções de índice de compatibilidade e índice de alerta foram definidas, sendo que os parâmetros que determinam o comportamento do aprendizado participativo foram discutidos.

O próximo capítulo utiliza a noção do aprendizado participativo para obter um algoritmo de agrupamento. Conhecido como agrupamento participativo, esse algoritmo servirá de base para o desenvolvimento de modelos fuzzy funcionais participativos evolutivos.

Capítulo 3

Convergência do Aprendizado Participativo

3.1 Introdução

O aprendizado participativo como proposto no capítulo anterior pode ser visto como uma sequência cuja convergência é analisada nesse capítulo. Conhecida a natureza e origem das observações, busca-se verificar como o aprendizado participativo desenvolve as iterações. Caso os dados possam ser representados por funções, este capítulo mostra algumas características e analogias que podem ser propostas. Caso os dados sejam uma variável aleatória, pode-se obter o valor esperado do aprendizado participativo ao longo do espaço.

Numericamente, a sequência $\{v^k\}$ do aprendizado participativo como definida no Capítulo 2 pode ser vista como um procedimento recursivo e iterativo. Dessa forma, a sequência $\{v^k\}$ representa uma busca local, e sua convergência está relacionada com a existência de determinadas singularidades no espaço das entradas x^k .

No caso particular, no qual as entradas x^k são representadas como uma função de v^k , a construção da sequência $\{v^k\}$ é tal que visa encontrar regiões que contenham um zero local, pontos de mínimo ou máximo, ou mesmo um ponto fixo relacionado à essa função que descreve x^k . Nesse caso, encontrar um ótimo local ou um ponto fixo pode ser visto também como um problema de busca por um zero dessa função (Conte e de Boor

1981).

Em contrapartida, numa situação onde x^k não represente necessariamente uma função conhecida, cada entrada pode ser vista como a realização de uma variável aleatória desconhecida. Nesse caso, o comportamento da sequência $\{v^k\}$ definida no aprendizado participativo depende do valor médio da variável aleatória x^k (Wasan 1969).

Nas seções seguintes, serão descritos esse caso particular no qual x^k é uma função conhecida e o caso mais geral no qual x^k é uma variável aleatória. A idéia é elaborar uma abordagem preliminar mais precisa acerca do comportamento matemático do aprendizado participativo no espaço dos dados. Não há uma literatura concisa sobre esse assunto mas, de acordo com algumas analogias que serão mostradas, o aprendizado participativo pode ser visto como um mecanismo linear direcionado, em cada instante, às observações x^k que estão sendo gradativamente conhecidas pelo modelo.

3.2 Caso Particular

Suponha que as entradas x^k sejam tais que exista uma função $f : [0, 1] \rightarrow [0, 1]$ contínua onde $x^k = f(v^k)$. Com isso, o processo iterativo associado ao aprendizado participativo mostrado no Capítulo 2 é reescrito como:

$$v^{k+1} = v^k + \alpha^k G^k(f(v^k) - v^k) \quad (3.1)$$

Assim, pela continuidade de f em $[0, 1]$, existe pelo menos um ponto $\bar{v} \in [0, 1]$ no qual $f(\bar{v}) = \bar{v}$, o qual é chamado de ponto fixo da função f (Lages 2003). Com efeito, definindo $h(v) = f(v) - v$, tem-se que $h : [0, 1] \rightarrow [-1, 1]$ também é contínua e, como a origem $0 \in [-1, 1]$, pelo Teorema do Valor Intermediário existe um $\bar{v} \in [0, 1]$ tal que $h(\bar{v}) = 0$ (Lages 2003).

Tem-se assim o seguinte resultado acerca do aprendizado participativo como

descrito em (3.1):

Proposição 3.1. *Seja $f : [0, 1] \rightarrow [0, 1]$ uma função contínua e $I \subseteq [0, 1]$ um intervalo real no qual $f'(v) = \frac{df(v)}{dv}$ também seja contínua para $\bar{v} \in I$. Se $f'(v) < 0$ em I , então para $\alpha^k \in (0, 1]$ tem-se $\lim_{k \rightarrow \infty} v^k = \bar{v}$, isto é, a sequência definida em (3.1) converge para o ponto fixo de f , caso exista.*

Demonstração. Tomando o valor inicial $v^1 \in I$, tem-se duas possibilidades (excetuando-se a óbvia onde $v^1 = \bar{v}$):

(i) $f(v^1) - v^1 > 0$;

(ii) $f(v^1) - v^1 < 0$.

(i): No primeiro caso, no qual $f(v^1) - v^1 > 0$, para $\alpha^k G^k \in [0, 1]$ tem-se:

$$k = 1 : \quad v^2 = v^1 + \alpha^1 G^1 (f(v^1) - v^1) > v^1; \text{ (lembrando que } f(v^1) \in [0, 1])$$

$$k = 2 : \quad v^3 = v^2 + \alpha^2 G^2 (f(v^2) - v^2) > v^2;$$

$$\vdots \quad v^{k+1} > v^k$$

afirmando, como é provado em seguida e ilustrado na Figura 3.1, que $(f(v^k) - v^k)$ não muda de sinal nesse intervalo.

Assim, tem-se que $\{v^k\}$ é uma sequência monótona crescente e limitada, portanto existe um $\bar{v}$ tal que $\lim_{k \rightarrow \infty} v^k = \bar{v}$ (Lages 2003).

Sabendo que existe o limite de $\{v^k\}$, para mostrar que $f(\bar{v}) = \bar{v}$, ou seja, este limite é o ponto fixo, basta notar pela expressão (3.1) que:

$$\lim_{k \rightarrow \infty} v^{k+1} = \lim_{k \rightarrow \infty} v^k + \lim_{k \rightarrow \infty} \alpha^k G^k (f(v^k) - v^k) \Rightarrow \lim_{k \rightarrow \infty} \alpha^k G^k (f(v^k) - v^k) = 0$$

e, como $\alpha^k G^k \in [0, 1]$ é, por definição, finito então: $\lim_{k \rightarrow \infty} (f(v^k) - v^k) = 0$.

Dessa forma, pela continuidade de f , tem-se:

$$f(\lim_{k \rightarrow \infty} v^k) = \lim_{k \rightarrow \infty} v^k \Rightarrow f(\bar{v}) = \bar{v}$$

Com isso, nessa situação $\bar{v} = \sup_{k \in \mathbb{N}} \{v^k\}$, ou seja, o ponto fixo é a cota superior da sequência $\{v^k\}$ (Lages 2003).

Logo, falta apenas mostrar a afirmação que $(f(v^k) - v^k)$ não muda de sinal no intervalo I . De fato, suponha que exista um instante j no qual $(f(v^j) - v^j) > 0$ e $(f(v^{j+1}) - v^{j+1}) < 0$, isto é, um instante no qual a sequência $\{v^k\}$ mude de sinal. Portanto:

$$f(v^j) - f(v^{j+1}) - v^j + v^{j+1} > 0 \quad \Rightarrow \quad \frac{f(v^j) - f(v^{j+1})}{v^j - v^{j+1}} > 1 > 0$$

Pelo Teorema do Valor Médio (Lages 2003), existe um $v^* \in (v^j, v^{j+1})$ onde $f'(v^*) > 1 > 0$, o que contradiz a hipótese inicial de que $f'(\bar{v}) < 0$, pois o ponto fixo $\bar{v} \in (v^j, v^{j+1})$. Logo, pode-se afirmar que $(f(v^k) - v^k)$ não muda de sinal para todo $k = 1, 2, \dots$.

(ii): No caso onde $f(v^1) - v^1 < 0$ a prova é idêntica, bastando notar, pelo mesmo raciocínio, que $\{v^k\}$ seria uma sequência monótona decrescente e limitada e que o ponto fixo $\bar{v} = \inf_{k \in \mathbb{N}} \{v^k\}$, ou seja, o ponto fixo seria a cota inferior da sequência $\{v^k\}$, o que completa a demonstração. $\square$

No entanto, pela Proposição 3.1, caso exista o ponto fixo $\bar{v}$ mas $f'(\bar{v}) > 0$, não há garantia de convergência para sequência do aprendizado participativo $\{v^k\}$. Dessa forma, a proposição seguinte define as condições de convergência quando $f'(\bar{v}) > 0$.

Proposição 3.2. *Sejam as hipóteses da Proposição 3.1 com $f'(v) > 0$ no intervalo I . Então, para $\alpha^k \in [-1, 0)$, a sequência do aprendizado participativo $\{v^k\}$ é tal que $\lim_{k \rightarrow \infty} v^k = \bar{v}$.*

Demonstração. Para a prova basta seguir os mesmos passos da demonstração da Proposição 3.1, apenas observando que nesse caso $\{v^k\}$ é monótona crescente se $f(v^1) - v^1 < 0$ e monótona decrescente se $f(v^1) - v^1 > 0$. $\square$

Logo, tem-se o seguinte corolário que garante a convergência do aprendizado participativo $\{v^k\}$ considerado na expressão (3.1) sempre que existir um ponto fixo.

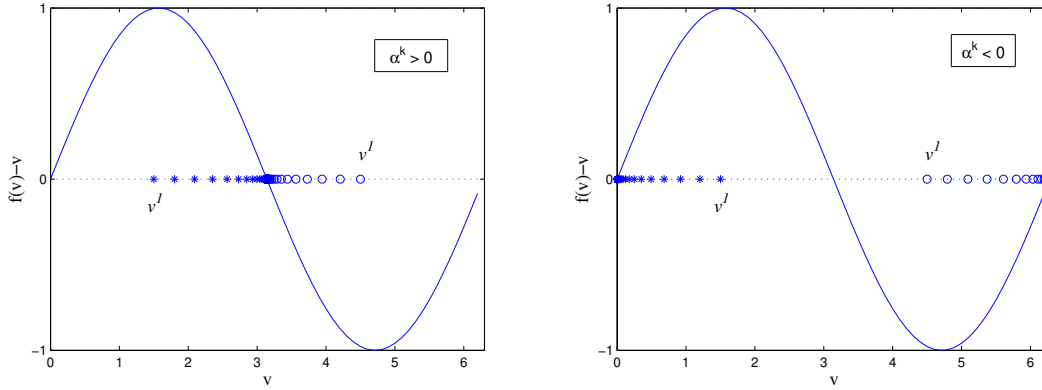


Figura 3.2: Exemplo com $f(v) = \sin(v) + v$ e $v^{k+1} = v^k + \alpha G^k \sin(v^k)$.

Com isso, para qualquer f contínua e diferenciável existem condições que garantem a convergência da sequência $\{v^k\}$ definida em (3.1). Para obter a taxa de convergência de $\{v^k\}$, isto é, a velocidade na qual $v^k \rightarrow \bar{v}$, considere a seguinte definição (Conte e de Boor 1981):

Definição 3.1. Uma sequência $\{v^k\}$ converge para um valor $\bar{v}$ em ordem $q > 0$, se existe uma constante real positiva K tal que:

$$|v^{k+1} - \bar{v}| \leq K |v^k - \bar{v}|^q$$

Portanto, tem-se a seguinte proposição:

Proposição 3.3. Considere as hipóteses do Corolário 3.1. Se existe $\lim_{k \rightarrow \infty} v^k = \bar{v}$ então, numa vizinhança de $\bar{v}$, $\{v^k\}$ converge em ordem $q = 1$ (converge linearmente).

Demonstração. Considere a função $h(v) = (f(v) - v)$, $\theta^k = (v^k - \bar{v})$ e $\alpha^k G^k \in [0, 1]$. Logo, utilizando a expressão (3.1) para $\{v^k\}$ tem-se:

$$v^{k+1} = v^k + \alpha^k G^k h(v^k) \quad \Rightarrow \quad \theta^{k+1} = \theta^k + \alpha^k G^k h(v^k) \leq \theta^k + h(v^k)$$

e, deve-se mostrar que existe uma constante real positiva K na qual $|\theta^{k+1}| \leq K |\theta^k|$.

Para tal, basta observar que para todo v^k próximo $\bar{v}$, pelo Teorema do Valor Médio existe

um ponto ξ entre v^k e $\bar{v}$ onde:

$$h'(\xi) = \frac{h(v^k) - h(\bar{v})}{v^k - \bar{v}} = \frac{h(v^k) - 0}{\theta^k} \Rightarrow h(v^k) = h'(\xi)\theta^k$$

Portanto, tem-se $\theta^k + h(v^k) = [1 + h'(\xi)]\theta^k$ e, supondo $h'(\xi)$ limitado, existe uma constante positiva K onde $|h'(\xi)| < K$. Assim:

$$|\theta^{k+1}| \leq |[1 + h'(\xi)]\theta^k| \leq (1 + K)|\theta^k|$$

o que completa a demonstração. $\square$

Logo, $\{v^k\}$ converge em ordem $q = 1$, ou seja, a sequência $\{v^k\}$ converge linearmente para o ponto fixo $\bar{v}$, sempre que o limite existir (Conte e de Boor 1981). Em geral, casos como o definido em (3.1) convergem linearmente pois partem de informações de ordem zero de f , onde supõe-se conhecido apenas o valor da função em cada ponto.

Vale notar, que o Corolário 3.1 e a Proposição 3.3 são condições necessárias para a convergência linear de (3.1) e, em casos onde não existe a derivada de f ou mesmo quando f não é contínua, a sequência $\{v^k\}$ também pode convergir.

Situações mais gerais podem ser interpretadas a partir de uma abordagem onde $f(v^k)$ não necessariamente existe ou é desconhecida. Nesses casos, é razoável supor que x^k é a realização de uma variável aleatória com uma certa distribuição, e o comportamento convergente de v^k depende, portanto, da média ou esperança matemática dessa variável aleatória x^k , como será mostrado na seção seguinte.

3.3 Caso Geral

Em problemas mais complexos, é difícil identificar as funções que descrevam as observações x^k . Com isso, em detrimento do caso bem determinado descrito na se-

ção anterior, x^k pode ser visto como uma variável aleatória com distribuição desconhecida (Feller 1968). Tal premissa remete o aprendizado participativo descrito no Capítulo 2 à problemas semelhantes aos conhecidos em aproximação estocástica (Wasan 1969), (Kushner e Clark 1978).

Dentre as aplicações mais utilizadas para aproximação estocástica está a solução de problemas de otimização local ou zeros e raízes de funções. Se, para cada realização de uma variável aleatória, uma direção puder ser definida, então um estimador para métodos do tipo gradiente também pode ser obtido (Wang e Chong 1998). Baseados nisso, os primeiros algoritmos de aproximação estocástica foram desenvolvidos ainda na década de 50 (Robbins e Monro 1951), (Dvoretzky 1954).

A equação do aprendizado participativo definida em (2.1) é dada por:

$$v^{k+1} = v^k + \alpha^k G^k(x^k - v^k)$$

E possui uma interpretação associada ao contexto participativo dado pelos índices de compatibilidade (2.3) e alerta (2.4) considerados no termo G^k .

No entanto, considerando as observações x^k como realizações de uma variável aleatória, algumas analogias remetem à problemas de aproximação estocástica (Wasan 1969), (Kushner e Clark 1978) e alguns resultados conhecidos podem ser utilizados na análise do aprendizado participativo.

Inicialmente, considere num instante k a média dos valores x^k :

$$M^k = \frac{1}{k} \sum_{i=1}^k x^i \tag{3.2}$$

de onde tem-se que:

$$(k+1)M^{k+1} = kM^k + x^{k+1} \quad \Rightarrow \quad M^{k+1} = M^k + \frac{1}{k+1}(x^{k+1} - M^k) \tag{3.3}$$

Numericamente existe um valor para o produto $\alpha^k G^k$ no qual a sequência definida no aprendizado participativo é semelhante à média calculada recursivamente (3.3).

Considerando $v^{k+1} = v^k + \alpha^k G^k (x^k - v^k)$, para $\alpha^k = \frac{1}{k+1}$ e $G^k = 1$ para todo $k = 1, 2, \dots$, tem-se:

$$v^{k+1} = v^k + \frac{1}{k+1} (x^k - v^k) \quad (3.4)$$

definindo uma expressão similar à (3.3).

Porém, nesse caso G^k é tal que não depende de x^k e v^k e desconsidera o contexto participativo definido pelos índices de compatibilidade e alerta. Assintoticamente, para $k \rightarrow \infty$, espera-se que $v^{k+1} = v^k + \alpha^k (\rho^k)^{1-\alpha^k} (x^k - v^k)$ definido em (2.6) no Capítulo 2 convirja para o mesmo valor que a sequência (3.4), sendo assim a análise é feita a partir desse caso mais simples definido em (3.4).

Em (Wang, Chong e Kulkarni 1996) há uma discussão acerca de 6 condições equivalentes de convergência para modelos de aproximação estocástica. Dentre elas, é interessante destacar:

$$\lim_{k \rightarrow \infty} \alpha^k = 0 \quad \text{e} \quad \sum_{k=1}^{\infty} \alpha^k = \infty \quad (3.5)$$

Portanto, para considerar a sequência do aprendizado participativo também com o termo participativo em G^k , uma sugestão seria defini-lo de forma a manter o limite de α^k mostrado em (3.5), ou seja:

$$\lim_{k \rightarrow \infty} \alpha^k G^k = \lim_{k \rightarrow \infty} \alpha^k \quad (3.6)$$

Assim, analisar a convergência da sequência (2.6) mostrada no Capítulo 2 seria equivalente à analisar a convergência da sequência (3.4). Em outras palavras, os resultados conhecidos em métodos de aproximação estocástica sugerem que a sequência

do aprendizado participativo $\{v^k\}$ converge para a média de x^k . No entanto, o termo $\alpha^k G^k$ deve satisfazer as mesmas condições que α^k , isto é, $\alpha^k G^k > 0$ e $\alpha^k G^k \rightarrow 0$ quando $k \rightarrow \infty$ (Wang et al. 1996).

Claramente a média calculada recursivamente em (3.3) satisfaz as condições (3.5) e, de fato, a sequência $\{M^k\}$ é convergente. Com isso, considerando a simplificação da sequência do aprendizado participativo em (3.4), tem-se o seguinte resultado:

Proposição 3.4. *Sejam a sequência $\{M^k\}$ que define a média mostrada em (3.2) e a expressão para a sequência $\{v^k\}$ em (3.4). Com $x^k \in [0, 1]$ para todo $k = 1, 2, \dots$, se $\{M^k\}$ é convergente então $\{v^k\}$ também é convergente e $\lim_{k \rightarrow \infty} v^k = \lim_{k \rightarrow \infty} M^k$.*

Demonstração. Pela expressão de v^k em (3.4) tem-se que:

$$x^i = (i+1)v^{i+1} - iv^i$$

e, substituindo na expressão de M^k em (3.2):

$$M^k = \frac{1}{k} \sum_{i=1}^k x^i \Rightarrow M^k = \frac{1}{k} \sum_{i=1}^k [(i+1)v^{i+1} - iv^i]$$

Desenvolvendo a soma acima, e fácil notar que:

$$\sum_{i=1}^k [(i+1)v^{i+1} - iv^i] = 2v^2 - v^1 + 3v^3 - 2v^2 + 4v^4 - 3v^3 + 5v^5 - 4v^4 + \dots + (k+1)v^k = (k+1)v^k - v^1$$

portanto, sabendo que $\lim_{k \rightarrow \infty} \frac{k+1}{k} = 1$ e $\lim_{k \rightarrow \infty} \frac{1}{k} = 0$, e pela hipótese de existência do limite de M^k obtém-se que:

$$M^k = \frac{k+1}{k} v^{k+1} - \frac{1}{k} v^1 \Rightarrow \lim_{k \rightarrow \infty} v^{k+1} = \lim_{k \rightarrow \infty} M^k$$

Logo, $\{v^k\}$ converge para o mesmo valor da média M^k , desde que os limites existam, o que completa a demonstração. $\square$

Alternativamente, as variáveis aleatórias x^k podem também ser representadas como $x^k = \bar{v} + \xi$, ou seja, observações com um certo ruído ξ e uma componente determinística $\bar{v}$. Assim, pode-se também considerar o termo participativo G^k e analisar a seqüência $\{v^k\}$ como definido em (2.1) no Capítulo 2.

Suponha, sem perda de generalidade, que o ruído tenha média nula e, portanto, $E(x^k) = \bar{v}$, sendo $E(\cdot)$ o valor esperado (Feller 1968). Assim, considerando o termo α^k satisfazendo (3.5) e o termo $\alpha^k G^k$ satisfazendo (3.6), a seqüência $\{v^k\}$ converge em média quadrática para $\bar{v}$ (Wang et al. 1996).

Em outras palavras, considerando a expressão do aprendizado participativo (2.1) mostrada no Capítulo 2 analogamente à um modelo de aproximação estocástica, as condições (3.5) e (3.6) garantem que existe $\lim_{k \rightarrow \infty} [E(v^k) - \bar{v}]^2 = 0$ e (Dvoretzky 1954), (Saridis 1974), (Wang et al. 1996):

$$\lim_{k \rightarrow \infty} [E(v^k) - \bar{v}]^2 = 0 \quad (3.7)$$

Portanto, $\{v^k\}$ converge para a média de x^k , considerando o termo participativo $G^k \in [0, 1]$ satisfazendo as condições para α^k em (3.5) e as condições para $\alpha^k G^k$ em (3.6).

3.4 Resumo

Esse capítulo buscou formalizar uma análise do aprendizado participativo. Os resultados obtidos acerca da convergência serão de grande importância para o estudo do comportamento do algoritmo de agrupamento que é apresentado no capítulo seguinte.

Vale notar que, não encontrando tais resultados na literatura vigente, buscou-se acender aqui as primeiras idéias sobre o estudo formal da noção de aprendizado proposta por (Yager 1990), de onde espera-se que sirva de ajuda para desenvolvimentos seguintes.

Capítulo 4

Agrupamento Participativo

4.1 Introdução

Neste trabalho o conceito de agrupamento utilizado assemelha-se aos algoritmos de agrupamento por partição, onde o objetivo é dividir um conjunto de dados em subconjuntos, e associar cada subconjunto à um ponto representativo chamado centro de grupo, ponto focal ou protótipo do grupo (Bezdek e Pal 1992).

O algoritmo Fuzzy *C-Means* (Bezdek 1981) e suas extensões (Bezdek, Tsao e Pal 1992), (Kaymak e Setnes 2000) são exemplos de agrupamento por partição. No Fuzzy *C-Means* (FCM), o agrupamento é formulado como um problema de otimização cuja solução são os centros dos grupos e os respectivos graus de pertinência de cada dado.

No agrupamento participativo proposto em (Silva 2003) e discutido nesse capítulo, a componente participativa definida pelos índices de compatibilidade e alerta estipula o número de grupos, isto é, estipula o número de subconjuntos utilizados para representar o conjunto de dados. O objeto de aprendizagem é a estrutura de grupo, estrutura esta representada pelo conjunto dos centros de grupo.

Em geral, encontrar o número ótimo de grupos para representar eficientemente determinado conjunto de dados exige a definição de uma função de validação, a ser utilizada como critério de otimalidade. Porém, mesmo definindo a função de validação, o processo de agrupamento não é único e depende fortemente do conjunto de dados, das escolhas dos parâmetros do algoritmo de agrupamento, da função de validação definida e

ainda do problema em questão (Roubens 1982), (Bezdek e Pal 1992), (Geva, Steinberg, Bruckmair e Nahum 2000).

Em um problema de agrupamento, quando o número de grupos é um parâmetro de entrada fixo a ser determinado, o algoritmo de agrupamento é dito supervisionado. Por outro lado, um algoritmo capaz de definir o número de grupos de acordo com os dados de entrada é chamado não supervisionado (Gath e Geva 1989).

Algoritmos de agrupamento não supervisionados são importantes em aplicações onde a entrada dos dados é feita sequencialmente e a estrutura de grupo deve ser alterada em cada instante como, por exemplo, em problemas de previsão de séries temporais em tempo real (Lima, Gomide e Ballini 2006). Nesse caso, a estrutura inicial é estimada, mas os dados a serem agrupados são coletados continuamente. Portanto, o algoritmo precisa ser desenvolvido de forma a alterar, se necessário, o número de grupos em cada instante que um novo dado é obtido.

O aprendizado participativo pode ser utilizado para a construção de um algoritmo de agrupamento não supervisionado, chamado agrupamento participativo (Silva 2003), (Silva et al. 2005). Diferentemente dos algoritmos onde os centros são construídos com solução de um problema de otimização, no agrupamento participativo os centros são modificados gradativamente utilizando o conceito de aprendizagem participativa. Portanto, analisar o comportamento da sequência $\{v^k\}$ torna-se importante, como feito no Capítulo 3 deste trabalho.

4.2 Algoritmo de Agrupamento Participativo

No agrupamento participativo, os centros que representam os grupos são atualizados de acordo com o mecanismo do aprendizado participativo definido no Capítulo 2. Em outras palavras, se c^k é o número de grupos, o centro de cada grupo é definido

como $v_i^k \in [0, 1]^p$, onde $k = 1, 2, \dots$ são os instantes nos quais os dados são fornecidos e $i = 1, 2, \dots, c^k$. Pela expressão (2.6) do aprendizado participativo no Capítulo 2, a atualização de cada centro de grupo é dada por:

$$v_i^{k+1} = v_i^k + \alpha(\rho_i^k)^{1-a_i^k}(x^k - v_i^k) \quad (4.1)$$

considerando ρ_i^k o índice de compatibilidade da entrada x^k com o centro v_i^k e a_i^k o índice de alerta do grupo i nos instantes $k = 1, 2, \dots$.

Como mostrado em (Silva 2003), o algoritmo de agrupamento participativo utiliza os índices de compatibilidade ρ_i^k como critério para atualização dos grupos, e os índices de alerta a_i^k como critério para a criação de novos grupos. Um novo grupo é criado sempre que um limiar for ultrapassado pelos índices de alerta. Esse limiar exprime, portanto, a necessidade de revisão no número de grupos. Em contrapartida, o centro com maior índice de compatibilidade é atualizado sempre que o limiar do índice de alerta estiver satisfeito. O agrupamento participativo pode ser visto analogamente aos algoritmos de aprendizagem competitiva (Hertz, Krogh e Palmer 1991), onde o centro de grupo mais próximo é atualizado.

O agrupamento participativo contempla também uma fase de verificação de grupos redundantes, detectando grupos distintos que representam regiões semelhantes. A verificação de grupos redundantes é feita utilizando os mesmos conceitos do índice de compatibilidade. Dados dois centros de grupo v_i^k e v_j^k distintos, a partir da expressão (2.3) no Capítulo 2, a compatibilidade entres esse centros é dada por:

$$\rho_{ij}^k = \rho_{ij}^k(v_i^k, v_j^k) = 1 - d(v_i^k, v_j^k) \quad (4.2)$$

sendo que quanto mais próximo de 1 for $\rho_{ij}^k \in [0, 1]$ maior é a redundância destes centros.

A Figura 4.1 mostra o comportamento do índice de compatibilidade entre os

centros (expressão (4.2)) em função da distância $d(v_i^k, v_j^k)$. No algoritmo de agrupamento participativo, o parâmetro $\lambda \in [0, 1]$ é definido de forma que dois grupos i, j distintos são considerados redundantes se o respectivo índice de compatibilidade for maior ou igual a esse parâmetro, isto é, se $\rho_{ij}^k \geq \lambda$, o que implica $d(v_i^k, v_j^k) > (1 - \lambda)$.

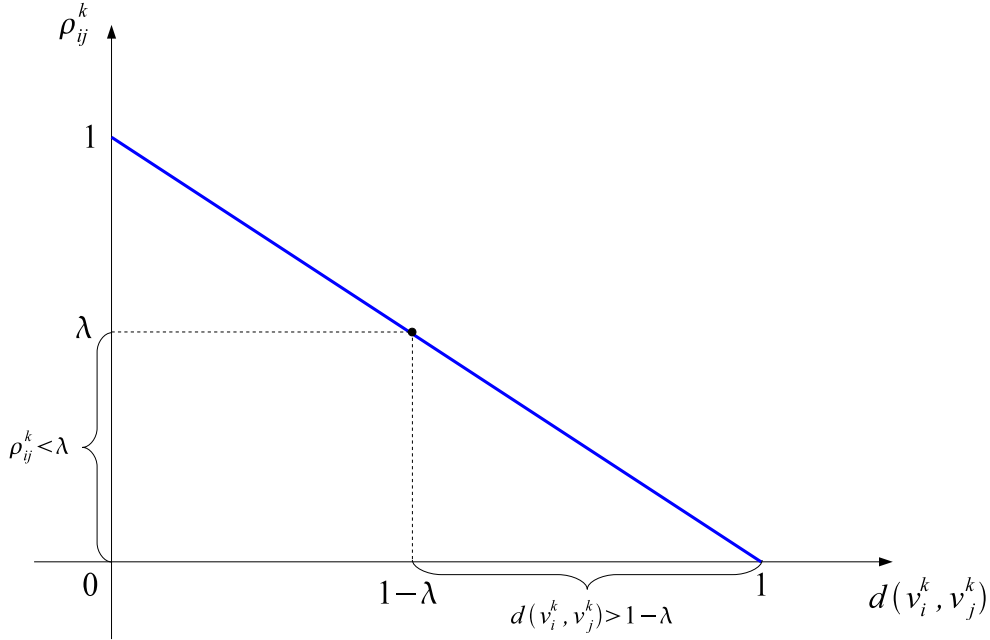


Figura 4.1: Índice de compatibilidade entre centros.

Para redefinir centros de grupo considerados redundantes, é feita uma combinação dos centros, obtida pela seguinte expressão:

$$v_r^k = v_i^k + \xi_{ij}^k(v_j^k - v_i^k) \quad (4.3)$$

sendo ξ_{ij}^k um valor real tal que $\xi_{ij}^k \in [0, 1]$.

Pela Figura 4.2, se $\xi_{ij}^k = 0$ ou $\xi_{ij}^k = 1$ na expressão (4.3) considera-se que um dos centros é excluído. Se $\xi_{ij}^k = 0.5$, o centro resultante v_r^k é definido como o ponto médio entre v_i^k e v_j^k . Uma outra possibilidade seria considerar $\xi_{ij}^k = \rho_{ij}^k$, o índice de compatibilidade entre centros definido em (4.2).

Utilizando, portanto, o aprendizado participativo como um mecanismo de

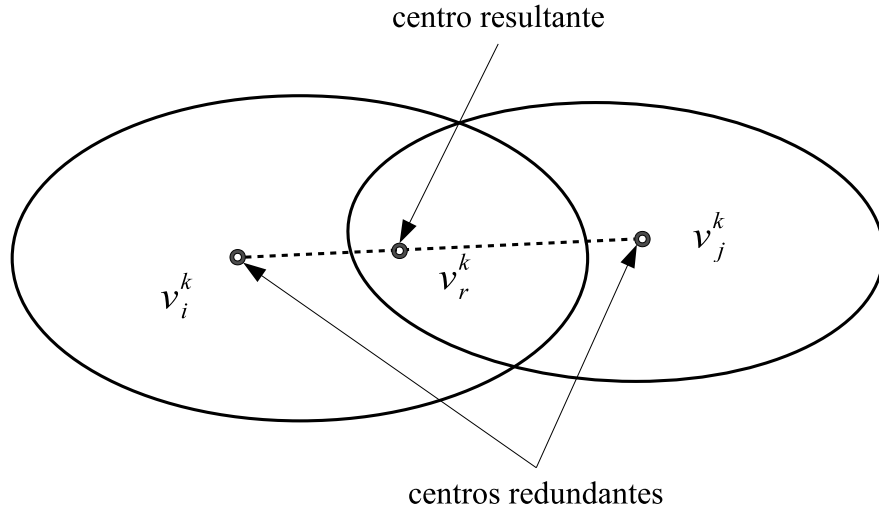


Figura 4.2: Redefinição dos grupos redundantes.

agrupamento, os passos do algoritmo do agrupamento participativo é detalhado a seguir. Nesse nova versão, diferentemente da definida em (Silva et al. 2005), não há critério de parada definido, sendo os centros obtidos de acordo com e entrada dos dados.

ALGORITMO AGRUPAMENTO PARTICIPATIVO

Inicializar o número de grupos $c^1 \geq 2$ e os centros $v_i \in [0, 1]^p$, $i = 1, 2, \dots, c^1$.

Escolher os parâmetros $\alpha \in [0, 1]$, $\beta \in [0, 1]$, limiar $\tau \in [0, 1]$ e $\lambda \in [0, 1]$.

Para $x^k \in [0, 1]^p$, $k = 1, 2, \dots$

Para $i = 1, 2, \dots, c^k$

 Calcular compatibilidade: $\rho_i^k = 1 - d(v_i^k, x^k)$;

 Calcular alerta: $a_i^{k+1} = a_i^k + \beta(1 - \rho_i^k - a_i^k)$;

Fim Para

Se $a_i^{k+1} \geq \tau$, $\forall i = 1, 2, \dots, c^k$ **então**

$c^k = c^k + 1$;

x^k é centro de um novo grupo: $v_{c^k}^k = x^k$;

Senão

 Obter o grupo mais compatível $s = \arg \max_{j=1, \dots, c^k} \{\rho_j^k\}$;

 Atualizar o grupo s : $v_s^{k+1} = v_s^k + \alpha(\rho_s^k)^{1-a_s^{k+1}}(x^k - v_s^k)$;

Fim Se

Para $i = 1, 2, \dots, (c^k - 1)$ e $j = (i + 1), \dots, c^k$
 Calcular compatibilidade entre centros v_i e v_j : $\rho_{ij}^k = 1 - d(v_i^k, v_j^k)$;
Se $\rho_{ij}^k \geq \lambda$ para algum j **então**
 excluir v_i^k e redefinir v_j^k usando os grupos i e j ;
 $c^k = c^k - 1$;
Fim Se
Fim Para
 $c^{k+1} = c^k$;

No algoritmo, a inicialização dos centros pode ser feita definindo os primeiros c^1 dados como centros de grupo, e os respectivos índices de alerta são inicializados como $a_i^0 = 0$ para todos os grupos, ou seja, o sistema é iniciado totalmente propenso a aprender.

Em cada k tem-se c^k grupos distintos representados pelos centros $v_j^k \in [0, 1]^p$, $j = 1, 2, \dots, c^k$ e atualizados utilizando a expressão (4.1). De acordo com o limiar τ , o número de grupos c^k pode aumentar, caso o índice de alerta da observação seja maior que o limiar em todos os grupos, permanecer o mesmo, caso o grupo de maior compatibilidade seja atualizado, ou ainda diminuir, caso dois grupos estejam próximos o suficiente.

Vale notar que o agrupamento participativo é não supervisionado e, para cada novo dado de entrada, o comportamento do algoritmo é definido pela taxa de aprendizagem α , pela sensibilidade β do índice de alerta (2.4) e ainda pelo limiar do índice de alerta τ e limiar do índice de compatibilidade entre centros λ . Enquanto α e β definem como os centros serão modificados, τ e λ estão associados ao tamanho (raio de ação) e ao número de grupos.

Considerações sobre os parâmetros definidos no agrupamento participativo são feitas na Seção 4.3.

4.3 Parâmetros do Agrupamento Participativo

Os parâmetros do algoritmo de agrupamento participativo são responsáveis pela forma como os dados são interpretados e os centros são atualizados. Nesta seção define-se algumas condições para α , β , τ e λ de maneira a facilitar a escolha dos respectivos valores, mantendo uma formação adequada da estrutura de grupo.

No agrupamento participativo, caso um dado provoque um aumento do índice de alerta acima do limiar $\tau \in [0, 1]$, esse dado será o centro de um novo grupo. Consequentemente, se em um instante o grupo de maior compatibilidade, s , for atualizado, então $a_s^k < \tau$, e pela expressão (2.4) de a_i^k tem-se:

$$\begin{aligned} a_s^k + \beta(1 - \rho_s^k - a_s^k) &= (1 - \beta)a_s^k + \beta d_s^k < \tau \\ \Rightarrow d_s^k &< \frac{\tau - (1 - \beta)a_s^k}{\beta} < \frac{\tau}{\beta} \end{aligned} \quad (4.4)$$

no qual $d_s^k = d(v_s^k, x^k)$ é a distância entre o centro do grupo s e o dado x^k no instante k .

Por outro lado, a partir do cálculo da compatibilidade entre os centros ρ_{ij}^k em (4.2) e da Figura 4.1, para quaisquer dois grupos $i, j = 1, 2, \dots, c^k$ distintos, tem-se:

$$\rho_{ij}^k = 1 - d_{ij}^k < \lambda \quad \Rightarrow \quad d_{ij}^k > 1 - \lambda \quad (4.5)$$

caso contrário, os grupos i e j seriam considerados redundantes e redefinidos como um único, sendo $d_{ij}^k = d(v_i^k, v_j^k)$ a distância entre os centros v_i^k e v_j^k .

Assim, pelo majorante (4.4) e pelo minorante (4.5) pode-se extrair uma relação entre os parâmetros β , τ e λ de forma a garantir uma coerência no processo de agrupamento. Uma sugestão é garantir que, sempre que um novo grupo for adicionado, o mesmo não seja visto como redundante no mesmo instante. Para isso, utilizando as

expressões (4.4) e (4.5), é interessante considerar que:

$$0 < \frac{\tau}{\beta} \leq 1 - \lambda \leq 1 \quad (4.6)$$

de onde tem-se:

$$\tau \leq \beta(1 - \lambda) \quad (4.7)$$

Na Figura 4.3 é ilustrado a condição (4.6) com um caso onde $d(v_s^k, x^k) > d(v_s^k, v_j^k)$ e x^k é mais propenso a definir um novo centro de grupo ao invés de atualizar algum centro existente.

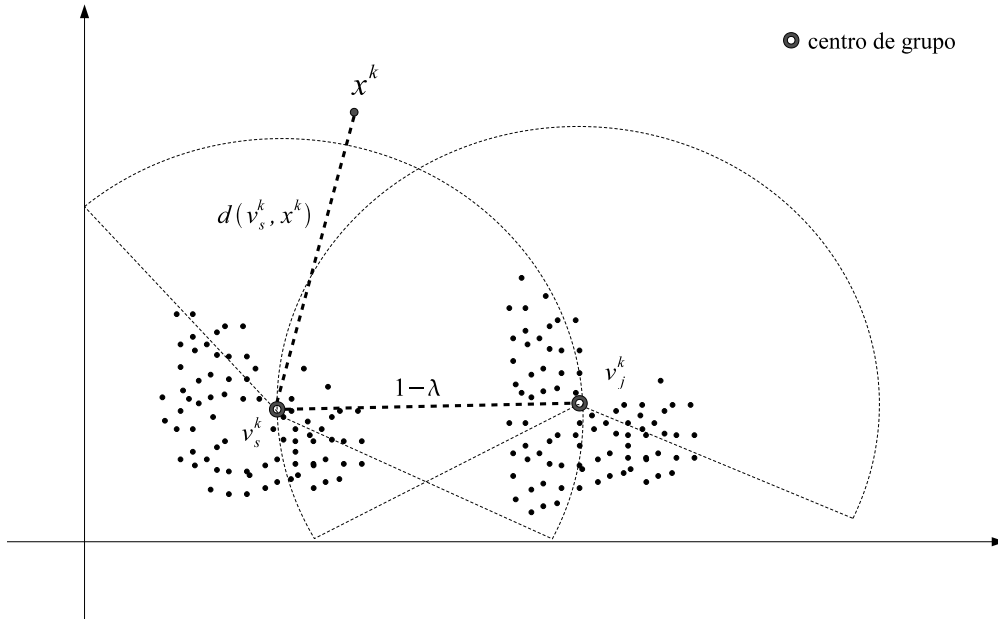


Figura 4.3: Caso onde um novo grupo não será redundante.

Em contrapartida, na Figura 4.4 é mostrada uma situação na qual $d(v_s^k, x^k) < d(v_s^k, v_j^k)$ e espera-se que o grupo mais compatível seja atualizado, tomando a condição (4.6) satisfeita. Nesse caso, como já ilustrado na Figura 4.1, se x^k fosse definido como centro de um novo grupo este seria excluído em seguida, o que na maioria das situações não é desejável.

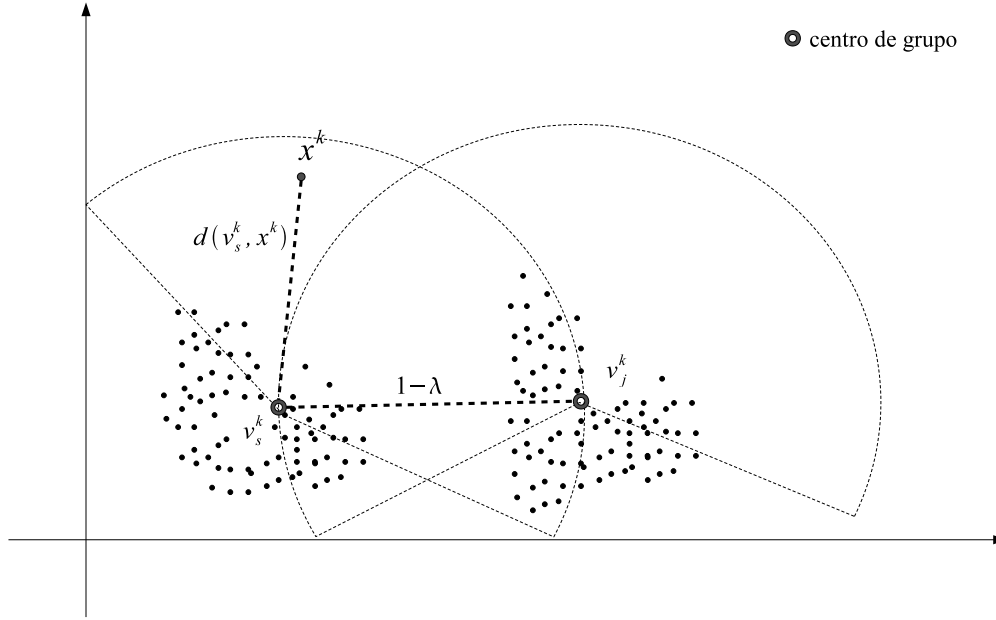


Figura 4.4: Caso onde é melhor atualizar um grupo existente.

O agrupamento participativo define uma estrutura de grupos representada pelos centros $v_i^k, i = 1, 2, \dots, c^k$, que em cada instante $k = 1, 2, \dots$ satisfazem pelo menos as duas propriedades seguintes:

$$\begin{aligned} a_i^k &\leq \tau && \text{para algum grupo } i \\ d_{ij}^k &\geq 1 - \lambda && \text{para quaisquer grupos } i, j \text{ distintos} \end{aligned} \quad (4.8)$$

com $i, j = 1, 2, \dots, c^k$, isto é, existe pelo menos um grupo cujo índice de alerta é menor que τ , e dois centros de grupo distintos distam entre si pelo menos $(1 - \lambda)$.

Observando também as análises detalhadas no Capítulo 3 sobre a convergência do aprendizado participativo, o agrupamento participativo pode ser visto como um método de aprendizagem semelhante aos modelos de aprendizagem competitiva (Hertz et al. 1991), (Kohonen 2001). Nesse contexto, apenas o grupo de maior índice de compatibilidade é atualizado, representando o vencedor no instante corrente. Deve-se notar que o agrupamento participativo é um modelo geral de agrupamento no qual, de acordo com certas condições, algumas características conhecidas podem ser extraídas.

Considere, por exemplo, $\beta = 0$. Nesse caso, usando (4.7) e considerando $\tau \in [0, 1]$ tem-se que $\tau = 0$. Logo, $a^{k+1} = a^k$ para todo $k = 1, 2, \dots$, e um grupo sempre é criado para qualquer valor de $\lambda \in [0, 1]$. Tomando $\lambda = 0$, em todo instante k um grupo é excluído mantendo, portanto, c^k constante, isto é, $c^k = c^*$ para todo $k = 1, 2, \dots$.

Valendo-se desse caso no qual c^k é constante e igual a c^* , cada centro de grupo $v_i^k, i = 1, 2, \dots, c^*$, será tal que convergirá para o valor médio dos dados com maior índice de compatibilidade no instante k . Partindo dos resultados mostrados no Capítulo 3, tem-se que:

$$\lim_{k \rightarrow \infty} [E(v^k) - \bar{v}]^2 = 0$$

sendo $\bar{v} = E(x^k)$ o valor médio das observações $x^k, k = 1, 2, \dots$.

Logo, nessas circunstâncias o agrupamento participativo torna-se o conhecido algoritmo de agrupamento *K-means*, no qual os centros dos grupos são a média das observações mais próximas (Duda e Hart 1973), isto é:

$$\bar{v}_i^k = \frac{1}{\sum_{j=1}^k n_i^j} \sum_{j=1}^k n_i^j x^j \quad \text{onde} \quad n_i^j = \begin{cases} 1, & \text{se } \rho_i^j = \max_{l=1, \dots, c^*} \{\rho_l^j\} \\ 0, & \text{caso contrário} \end{cases} \quad (4.9)$$

e portanto:

$$v_i^k \longrightarrow \bar{v}_i^k \quad (4.10)$$

ou seja, o centro de cada grupo converge para o valor médio dos dados de maior índice de compatibilidade, para $i = 1, 2, \dots, c^*$.

4.4 Resumo

Esse capítulo apresentou o algoritmo de agrupamento participativo, algoritmo derivado da noção de aprendizado participativo aplicada em problemas de agrupamento de dados. A característica essencial desse algoritmo é a forma como um novo dado é interpretado, podendo ocasionar uma atualização no aprendizado, ou um novo conhecimento. Em um caso, o número de grupos se mantém e algum centro de grupo é modificado, e no outro caso, um novo grupo é criado.

Detalhes acerca do comportamento do agrupamento participativo na determinação do número de grupos foram mostrados, definindo o agrupamento participativo como um algoritmo não supervisionado, cuja atualização dos centros é feita de acordo com índices de compatibilidade e índices de alerta.

O capítulo seguinte trata dos modelos fuzzy funcionais, enfatizando o modelo Takagi-Sugeno evolutivo. A utilização do agrupamento participativo na determinação desse tipo de modelo fuzzy funcional evolutivo serve de base para a proposta desse trabalho, que é definir o algoritmo do Aprendizado Participativo Evolutivo.

Capítulo 5

Modelos Fuzzy Funcionais Evolutivos

5.1 Introdução

Algoritmos de agrupamento como os mostrados no Capítulo 4 podem ser aplicados na construção do número de regras e dos antecedentes associados a cada uma das regras (Yen et al. 1998). Dessa forma, considera-se que o sistema possa ser descrito a partir de um conjunto de condições. Nos modelos fuzzy funcionais evolutivos esse conjunto de condições (regras) variam de acordo com o número de grupos obtidos.

Associando regras à cada grupo, os antecedentes das regras são representados pelos centros de grupo, e os conseqüentes são representados por ajustes lineares para cada regra, como ilustrado na Figura 5.1. A saída do modelo é calculada a partir da base de regras, via inferência fuzzy (Pedrycz e Gomide 1998). Caso o agrupamento seja feito a partir de um algoritmo não supervisionado e recursivo, a saída do modelo será obtida a partir da entrada atual e da base de regras atualizada recursivamente. Assim, tem-se um modelo evolutivo, cujos parâmetros são determinados continuamente e ajustados de forma gradativa (Angelov e Kasabov 2006).

A base de regras de um modelo fuzzy funcional associado à um agrupamento supervisionado possui um número fixo de regras e, geralmente, não são ajustadas recursivamente. Em cada instante os grupos são determinados utilizando todas as informações anteriores, independentemente dos grupos atuais (Magalhães 2004).

Por outro lado, se a base de regras de um modelo fuzzy funcional for ajustada

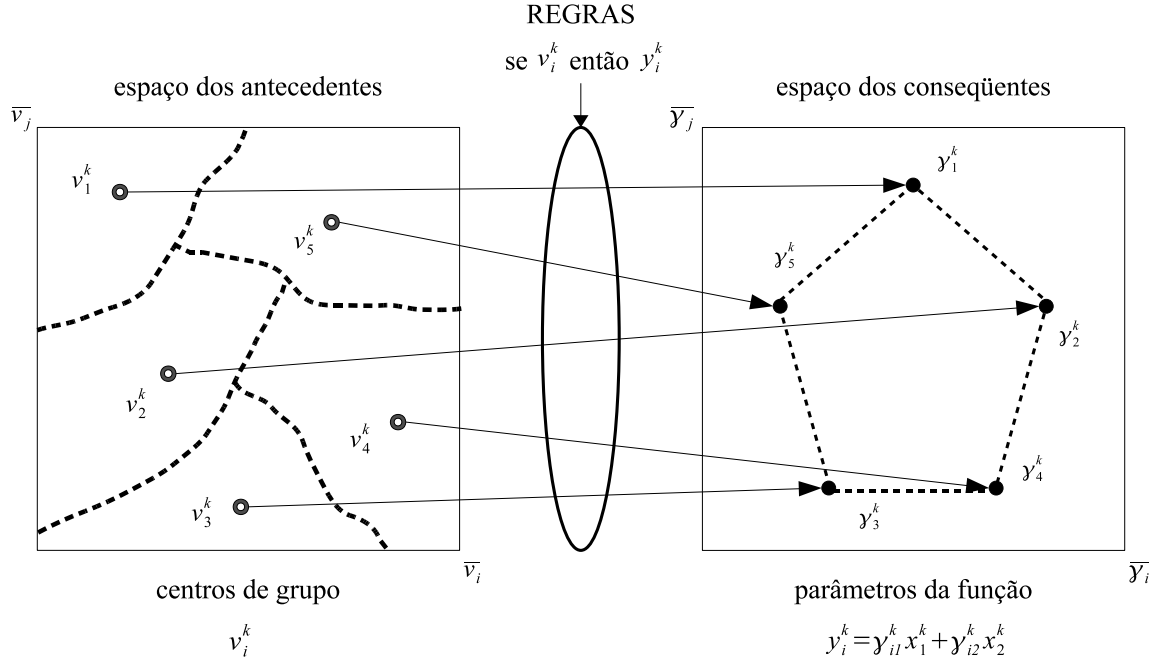


Figura 5.1: Espaço dos antecedentes e consequentes.

recursivamente e associada à um algoritmo de agrupamento não supervisionado, então o algoritmo pode determinar os grupos utilizando apenas o dado mais recente e os grupos atuais (Kasabov 2001), (Angelov e Filev 2004). Como vantagem, caso os dados sejam processados sequencialmente as atualizações feitas anteriormente nos centros são aproveitadas e, se necessário, o número de regras pode variar em cada instante que um novo dado é obtido. A modificação na base de regras é feita alterando também o número de regras ajustadas e recalculando os respectivos parâmetros do consequente. Pode-se adicionar novas regras que complementem a descrição dos dados observados, manter o número de regras atualizando alguma existente de acordo com a entrada, ou ainda diminuir o número de regras excluindo alguma que seja redundante.

Nos modelos fuzzy funcionais, os consequentes das regras são funções das variáveis de entrada. Os parâmetros destas funções podem ser ajustados recursivamente de acordo com os respectivos antecedentes utilizando, por exemplo, o método de mínimos quadrados recursivo (Young 1984).

5.2 Modelos Fuzzy Funcionais Takagi-Sugeno

Uma classe importante de modelos fuzzy funcionais são conhecidos como modelos Takagi-Sugeno (TS). Nesses modelos, os antecedentes das regras são termos lingüísticos (Klir e Yuan 1995) e os conseqüentes são funções das variáveis dos antecedentes (Takagi e Sugeno 1985). Esses tipos de modelos têm-se mostrado bastante eficientes na descrição de problemas não lineares e de natureza complexa.

De forma geral, em cada instante k no qual um dado x^k é obtido, um modelo fuzzy funcional considera regras na forma:

$$\mathcal{R}_i^k : \quad \text{SE } x^k \text{ é } A_i^k(x^k) \quad \text{ENTÃO } y_i^k = f_i^k(x^k) \quad (5.1)$$

$i = 1, 2, \dots, c^k$ e $k = 1, 2, \dots$, onde:

$\mathcal{R}_i^k$	i -ésima regra no instante k
$x^k \in [0, 1]^p$	vetor de entrada
y_i^k	saída da i -ésima regra
$A_i^k(x^k) \in [0, 1]^p$	vetor de funções de pertinência dos antecedentes
c^k	número de regras no instante k

Considerando que as regras descrevam comportamentos locais dos dados, o conseqüente de cada regra é definido por um modelo linear em cada um dos c^k grupos. A saída do modelo TS é, portanto, uma combinação das saídas de cada regra, y_i^k , ponderadas pelos respectivos graus de ativação. Dessa forma, tomando o centro de gravidade como mecanismo de inferência (Pedrycz e Gomide 1998), a saída y^k do modelo é dada por:

$$y^k = \frac{\sum_{j=1}^{c^k} \mu_j^k y_j^k}{\sum_{j=1}^{c^k} \mu_j^k} \quad (5.2)$$

O grau de ativação μ_i^k é definido como a interseção das funções de pertinência dos antecedentes e mede a representatividade da regra na composição da saída. Em $[0, 1]^p$, considere a expressão do grau de ativação dada por (Yen et al. 1998):

$$\mu_i^k = \mu_i^k(x^k) = A_{i1}^k(x^k)A_{i2}^k(x^k) \dots A_{ip}^k(x^k) = \prod_{j=1}^p A_{ij}^k \quad (5.3)$$

sendo que $A_i^k = A_i^k(x^k) = [A_{i1}^k \dots A_{ip}^k]$ é o vetor de funções de pertinência, $p \in \mathbb{N}$ é a dimensão do vetor de entrada x^k e $k = 1, 2, \dots$.

Modelos TS são de grande importância devido à possibilidade de descrição de sistemas não lineares e complexos a partir de um número reduzido de regras. A saída da base de regras y^k em (5.2) pode ser obtida ajustando os parâmetros a partir de um algoritmo de aprendizado (Yen et al. 1998).

Um modelo evolutivo é obtido associando uma regra do tipo (5.1) à cada região representada pelos grupos. Dentre os modelos que utilizam abordagens evolutivas, destacam-se os Modelo Takagi-Sugeno Evolutivo (*Evolving Takagi-Sugeno*), cuja interpretação dos dados é baseada na atualização recursiva de uma função potencial, calculada para a nova amostra dos dados e para os respectivos centros de grupo (Angelov et al. 2004). O algoritmo desenvolve a relação entre os potenciais do dado de entrada e dos centros de grupo, atualizando o número de regras e os respectivos parâmetros que a descrevem.

5.3 Modelo Takagi-Sugeno Evolutivo

Baseado no modelo TS, o Modelo Takagi-Sugeno Evolutivo (eTS) inclui um processo que atualiza o sistema de forma recursiva. Determinando os antecedentes das regras a partir de um agrupamento não supervisionado, o algoritmo eTS é desenvolvido adicionando-se novas regras ou modificando alguma regra existente no modelo, como esquematizado na Figura 5.2. Em cada instante, deve-se decidir como a base de regras será

atualizada, ajustando os parâmetros de acordo com o dado mais recente (Angelov 2002). O eTS baseia-se em uma noção onde a representatividade de um dado é medida a partir de uma função que define o potencial relativo à cada centro de grupo (Yager e Filev 1994).

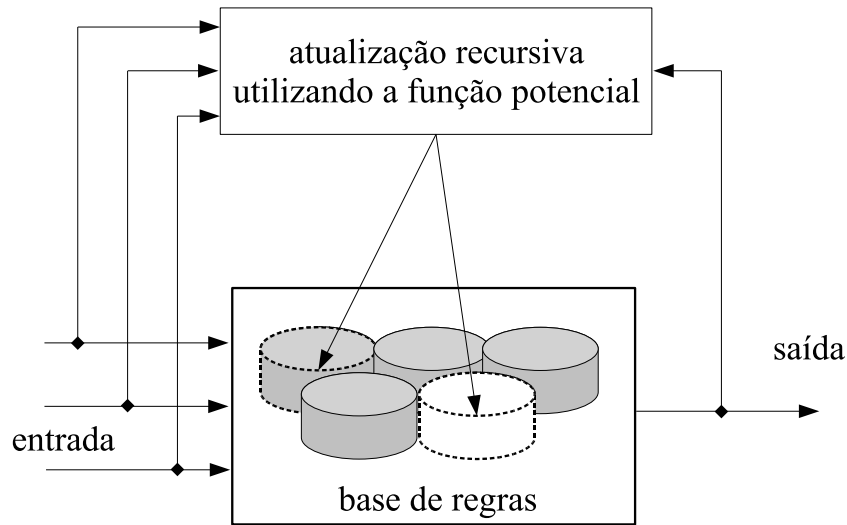


Figura 5.2: Mecanismo da base de regras no eTS.

Na forma recursiva, caso o potencial atualizado do novo dado seja maior que o potencial dos centros dos grupos existentes, então esse dado será um novo centro de grupo, podendo substituir algum centro existente ou representar o centro de um novo grupo. Após a fase de inicialização, o agrupamento é feito em cada instante, mantendo os centros dos grupos, atualizando algum centro existente ou criando um novo grupo (Angelov 2002). Os parâmetros da função linear do conseqüente são calculados utilizando o algoritmo de mínimos quadrados recursivo (Young 1984), que será descrito no próximo capítulo.

Assim, cada nova observação é vista como um possível centro de grupo, e quanto menor a distância em relação à algum centro existente, maior será o potencial dessa observação. A função potencial de um dado x^k é definida como (Angelov e Filev 2004):

$$P^k = P(x^k) = \frac{1}{k-1} \sum_{i=1}^{k-1} e^{-r\|x^k - x^i\|^2} \quad ; \quad k = 2, 3, \dots \quad (5.4)$$

onde r é uma constante positiva e $\|\cdot\|$ é a norma Euclidiana.

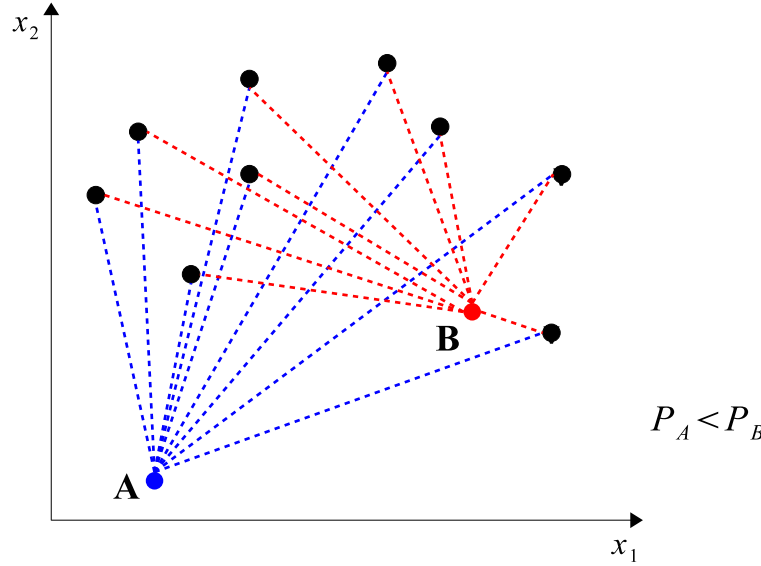


Figura 5.3: Ilustração da idéia de função potencial.

Como ilustrado na Figura 5.3, a função potencial busca detectar e quantificar regiões que possuam uma certa concentração de dados. Nessa figura, o potencial P_B calculado no ponto B é maior que o potencial P_A calculado no ponto A , pois as distâncias (ou a soma das distâncias) dos dados em relação ao ponto B são menores. Quanto mais concentrado estiverem os dados em uma certa região, maior será o potencial dos pontos dessa região. Intuitivamente, essa concentração de dados é caracterizada por pontos próximos entre si e, portanto, o potencial é uma função monótona e inversamente proporcional às distâncias entre as observações (Angelov e Kasabov 2006).

Análogo à esse conceito, uma função potencial também pode ser vista como uma medida de esparsidade dos dados (Pal e Chakraborty 2000). No eTS, os centros de grupo no instante seguinte são obtidos pela comparação entre os potenciais dos centros atuais e o potencial do dado mais recente.

Uma alternativa à função potencial definida em (5.4) é obtida a partir da função de Cauchy dada por (Angelov e Filev 2004):

$$P^k = \left(1 + \frac{\sum_{i=1}^{k-1} \|x^k - x^i\|^2}{k-1} \right)^{-1} \quad (5.5)$$

sendo que P^k denota o potencial do ponto x^k no instante $k = 2, 3, \dots$. Vale observar que P^k é inversamente proporcional à média das distâncias entre os dados conhecidos.

Em (Angelov e Zhou 2006) há a seguinte sugestão para o cálculo do potencial:

$$P^k = \left[1 + \vartheta^k + \frac{(\sigma^k - 2\varphi^k)}{(k-1)} \right]^{-1} \quad (5.6)$$

no qual P^k denota o potencial da observação x^k , $\vartheta^k = \|x^k\|^2$, $\sigma^k = \sum_{j=1}^{k-1} \|x^j\|^2$, $\varphi^k = \sum_{j=1}^{k-1} (x^j)^T x^k$, sendo $\|\cdot\|$ a norma Euclidiana, $(x^j)^T$ o transposto do vetor x^j e $k = 1, 2, \dots$.

Utilizando o potencial calculado para os centros dos grupos, a modificação das regras no eTS é feita de acordo com as seguintes condições (Angelov e Zhou 2006):

- a) se uma observação possuir o potencial maior que o potencial de todos os centros atuais, então esse ponto será um centro de grupo;
- b) se, além da condição anterior, o potencial do ponto for próximo o suficiente ao potencial de algum centro, então esse centro será substituído por essa observação;
- c) caso contrário, a base de regras permanece como está, sem modificação.

Em situações onde os dados são obtidos continuamente, o algoritmo de agrupamento deve decidir se um dado irá modificar um determinado grupo ou se um novo centro de grupo será definido. A cada novo dado x^k , os potenciais dos centros de grupo também são calculados e utilizados na decisão.

O processo de agrupamento envolvido no modelo eTS pode ser descrito de acordo com o seguinte algoritmo.

ALGORITMO AGRUPAMENTO DO ETS

Inicializar o número de grupos $c^1 \geq 2$ e os centros $v_i \in [0, 1]^p$, $i = 1, 2, \dots, c^1$.

Escolher o parâmetro $\delta > 1$.

Para $x^k \in [0, 1]^p$, $k = 1, 2, \dots$

Calcular o potencial (5.6) de x^k : $P^k = \left[1 + \vartheta^k + (\sigma^k - 2\varphi^k)/(k-1) \right]^{-1}$;

Para $i = 1, 2, \dots, c^k$

Calcular o potencial dos centros v_i^k : $P(v_i^k) = \frac{(k-1)P(v_i^{k-1})}{k-2 + P(v_i^{k-1})(1 + \|v_i^k - v_i^{k-1}\|^2)}$;

Fim Para

Se $P^k > P(v_i^k)$, $\forall i = 1, 2, \dots, c^k$ **então**

Se $\frac{P^k}{\max_{j=1, \dots, c^k} \{P(v_j^k)\}} \geq \delta$ **então**

x^k substitui o centro do grupo i : $v_i^{k+1} = x^k$;

$P(v_i^{k+1}) = P^k$;

Senão

$c^{k+1} = c^k + 1$;

x^k é centro de um novo grupo: $v_{c^{k+1}}^{k+1} = x^k$;

$P(v_{c^{k+1}}^{k+1}) = P^k$;

Fim Se

Fim Se

No algoritmo, o parâmetro $\delta > 1$ é obtido de acordo com a distância entre o centro atualizado e o centro mais próximo, portanto, depende do raio do grupo. Em (Angelov e Filev 2004) é sugerido $\delta = 1 + \frac{\delta_{min}}{r}$, onde δ_{min} é a distância entre o dado x^k e o centro de grupo mais próximo e r é o respectivo raio do grupo. Quanto maior for o valor de δ , menor será a chance da observação x^k substituir o centro de maior potencial, gerando provavelmente um número menor de grupos. A inicialização dos potenciais dos centros é obtida a partir da versão não recursiva da função P^k definida na expressão (5.5).

No conseqüente, os parâmetros são calculados de acordo com a alteração feita nos centros de grupo. Caso um novo grupo seja adicionado, os parâmetros da função linear

do respectivo conseqüente são obtidos a partir dos parâmetros das regras não modificadas. Por outro lado, caso algum grupo seja alterado, os parâmetros são atualizados utilizando mínimos quadrados recursivos (Young 1984), como mostrado no Capítulo 6.

A eficiência do eTS sugerida pela literatura serve de inspiração para a proposta desse trabalho, que será desenvolvida no Capítulo 6 aplicando o agrupamento participativo em um modelo fuzzy funcional.

5.4 Resumo

Esse capítulo apresentou os modelos fuzzy funcionais, especificamente o modelo Takagi-Sugeno. O agrupamento utilizado no algoritmo Takagi-Sugeno evolutivo foi detalhado, pois serve como base de comparação para o modelo proposto nesse trabalho.

O capítulo seguinte descreve o Aprendizado Participativo Evolutivo, que utiliza o agrupamento participativo na determinação de um modelo Takagi-Sugeno evolutivo.

Capítulo 6

Modelo Takagi-Sugeno Evolutivo Participativo

6.1 Introdução

O Aprendizado Participativo Evolutivo (APE) é um modelo fuzzy funcional evolutivo nos moldes do eTS descrito no Capítulo 5. No entanto, para a determinação dos antecedentes o APE utiliza o agrupamento participativo apresentado no Capítulo 4.

O APE utiliza a noção do aprendizado participativo para a interpretação dos dados e atualização dos antecedentes das regras. A base de regras é atualizada em cada instante, modificando o centro do grupo que possui maior índice de compatibilidade, ou criando uma nova regra caso o sistema possua alto índice de alerta e necessite de revisão. Em seguida, o APE determina os parâmetros da função do conseqüente utilizando o algoritmo de mínimos quadrados recursivo.

Unindo o mecanismo do aprendizado participativo descrito na Figura 2.1 ao mecanismo do algoritmo eTS mostrado na Figura 5.2, é mostrado na Figura 6.1 a base de regras no APE. O aprendizado participativo evolutivo cria uma nova regra ou modifica uma já existente. A partir do agrupamento participativo, se o índice de alerta for maior que um parâmetro $\tau \in [0, 1]$ escolhido, uma nova regra é criada, caso contrário, o centro de maior compatibilidade é atualizado seguindo a expressão do aprendizado participativo.

A partir de uma estrutura inicial de grupos, o APE determina as regras inici-

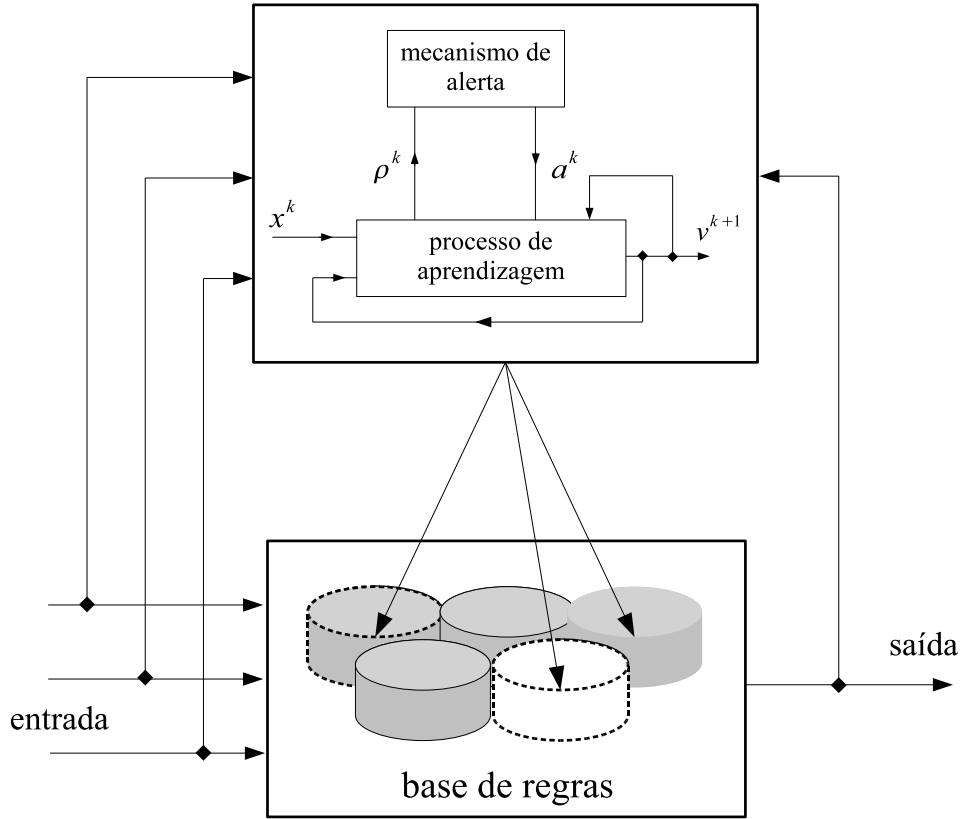


Figura 6.1: Mecanismo da base de regras no APE.

ais e a atualização nos antecedentes utilizando o agrupamento participativo. Associando os grupos obtidos à um modelo Takagi-Sugeno evolutivo, a modificação nos centros dos grupos implica alterar também os parâmetros do conseqüente da respectiva regra.

6.2 Aprendizado Participativo Evolutivo

O aprendizado participativo evolutivo (APE) considerado neste trabalho adota um modelo Takagi-Sugeno evolutivo construído de forma a ser modificado gradativamente de acordo com cada novo dado x^k . O APE atualiza os centros dos grupos utilizando o agrupamento participativo e ajusta nos antecedentes funções de pertinência Gaussianas com valor modal nesses centros, sendo que os parâmetros dos conseqüentes são obtidos por mínimos quadrados recursivos, como detalhado na Seção 6.2.2.

6.2.1 Atualização dos antecedentes das regras

A partir de uma estrutura inicial, a modificação da base de regras é feita considerando uma medida de compatibilidade, $\rho_i^k \in [0, 1]$, e um índice de alerta, $a_i^k \in [0, 1]$, com as mesmas propriedades definidas no Capítulo 4. Enquanto ρ_i^k mede o quanto o novo dado é compatível com a formação atual das regras, a_i^k representa uma medida crítica capaz de detectar o grau com que novas informações são coerentes com o modelo atual. O alerta é capaz de medir, portanto, a necessidade de revisão no modelo e adaptação da base de regras.

Além disso, o aprendizado participativo evolutivo possui um mecanismo que indica a presença de regras redundantes. Nesse caso, centros muito similares detectados no agrupamento participativo serão redefinidos para representar uma única regra com o objetivo de aumentar a eficiência do algoritmo e descrever o sistema com um número reduzido de regras.

Como comentado anteriormente, tomando o centro do grupo i no instante k , $v_i^k \in [0, 1]^p$, o objetivo do mecanismo de aprendizagem é atualizar essa variável a partir das observações reais x^k obtidas em cada instante.

Dado o valor do parâmetro $\tau \in [0, 1]$, que define um limiar para o índice de alerta a_i^k , uma regra é criada caso $a_i^k > \tau$ para todos os grupos. Caso contrário, a regra que possui o maior índice de compatibilidade é atualizada de acordo com a expressão:

$$v_i^{k+1} = v_i^k + G_i^k(x^k - v_i^k) \quad (6.1)$$

sendo que o termo $G_i^k \in [0, 1]$ é dado por:

$$G_i^k = \alpha(\rho_i^k)^{1-a_i^k}; \quad \rho_i^k = 1 - d_i^k \quad (6.2)$$

com $\alpha \in [0, 1]$ a taxa de aprendizagem, $d_i^k = d(x^k, v_i^k)$ a distância entre a observação x^k e o

centro v_i^k de maior índice de compatibilidade, ou seja:

$$i = \arg \max_j \{\rho_j^k\} \quad (6.3)$$

Seguindo a idéia do agrupamento participativo descrito no Capítulo 4, o índice de alerta a_i^k é dado por:

$$a_i^{k+1} = a_i^k + \beta(1 - \rho_i^{k+1} - a_i^k) \quad (6.4)$$

sendo β um parâmetro real tal que $\beta \in [0, 1]$.

O parâmetro β controla a taxa de atualização do índice de alerta. Quanto mais próximo de 1 for β , maior é a sensibilidade do sistema em relação às oscilações nos dados. Portanto, o centro v_i^k é modificado na direção da nova observação x^k , sendo que o tamanho do passo depende tanto do índice de compatibilidade quanto do índice de alerta.

Com isso, $a_i^k = 0$ implica $G_i^k = \alpha \rho_i^k$, gerando um mecanismo sem alerta. Quanto maior for o alerta, menor será a influência do índice de compatibilidade na atualização, lembrando que, por construção do agrupamento participativo, uma regra é atualizada somente se existe um grupo i tal que $a_i^k < \tau$.

O algoritmo APE também possui um mecanismo que calcula a compatibilidade entre os centros, pois a atualização dada pelo aprendizado participativo pode gerar centros muito próximos e, conseqüentemente, saídas com grau de ativação semelhantes. Assim, após a atualização, um índice de compatibilidade entre os centros é calculado e utilizado para evitar a criação de regras redundantes. Tomando o valor do parâmetro $\lambda \in [0, 1]$ como limiar, caso esse índice de compatibilidade seja maior que esse parâmetro λ , as regras são consideradas redundantes e serão substituídas por uma única.

O cálculo da compatibilidade entre os centros é feito de forma similar ao

índice de compatibilidade, ou seja:

$$\rho_{ij}^k = \rho_{ij}^k(v_i^k - v_j^k) = 1 - d(v_i^k, v_j^k) \quad (6.5)$$

sendo que $d(v_i^k, v_j^k)$ é a distância entre os centros v_i^k e v_j^k . Caso a condição $\rho_{ij}^k \geq \lambda$ seja satisfeita, a regra i é declarada como redundante e o centro do respectivo grupo é alterado.

Nas situações onde há centros de grupos redundantes, os mesmo podem ser redefinidos substituindo os centros por um novo centro representado pela combinação linear entre os redundantes. Como descrito no agrupamento participativo, diversas abordagens são possíveis, como obter o novo centro a partir da média dos centros redundantes ou, simplesmente, manter um grupo e excluir o outro.

A Tabela 6.1 compara o método de agrupamento utilizado pelo APE com o utilizado pelo modelo eTS apresentado no Capítulo 5. Vale ressaltar que a versão do eTS corresponde à proposta em (Angelov e Zhou 2006).

Tabela 6.1: Comparação entre os métodos de agrupamento no APE e no eTS.

APE	ETS
ADICIONA UM GRUPO	
se $a_i^k > \tau, \forall i = 1, \dots, c^k$	se $P_k > P(v_i^k), \forall i = 1, \dots, c^k$
ATUALIZA UM GRUPO	
se existe $a_i^k < \tau$	se $P_k > P(v_i^k)$ e $\frac{P_k}{P(v_i^k)} > \delta$ $P(v_i^k) = \max_j \{P(v_j^k)\}$
ATUALIZAÇÃO	
$v_s^{k+1} = v_s^k + \alpha(\rho_s^k)^{1-a_s^{k+1}}(x^k - v_s^k)$ $s = \arg \max_j \{\rho_j^k\}$	$v_i^{k+1} = x^k$ $P(v_i^k) = \max_j \{P(v_j^k)\}$
REMOVE UM GRUPO	
se $\rho_{ij}^k > \lambda$	originalmente não remove

Após o processo de agrupamento, as funções de pertinência Gaussianas dos antecedentes são calculadas para obter o grau de ativação da regra. Como ilustrado na Figura 6.2, em cada grupo os antecedentes A_i^k são calculados com valor modal dado de acordo com o valor do parâmetro r , e o grau de ativação da regra μ_i^k é dado por:

$$\mu_i^k = \prod_{j=1}^p A_{ij}^k = \prod_{j=1}^p e^{-r(v_{ij}^k - x_j^k)^2} = e^{-r\|v_i^k - x^k\|^2} \quad (6.6)$$

no qual $\|\cdot\|$ é a norma Euclidiana e r um parâmetro da função Gaussiana, $i = 1, 2, \dots, c^k$.

Os conseqüentes da base de regras no APE são construídos criando um modelo linear para cada grupo. Os parâmetros desse modelo linear são obtidos pelo método de mínimos quadrados recursivo, como detalhado a seguir.

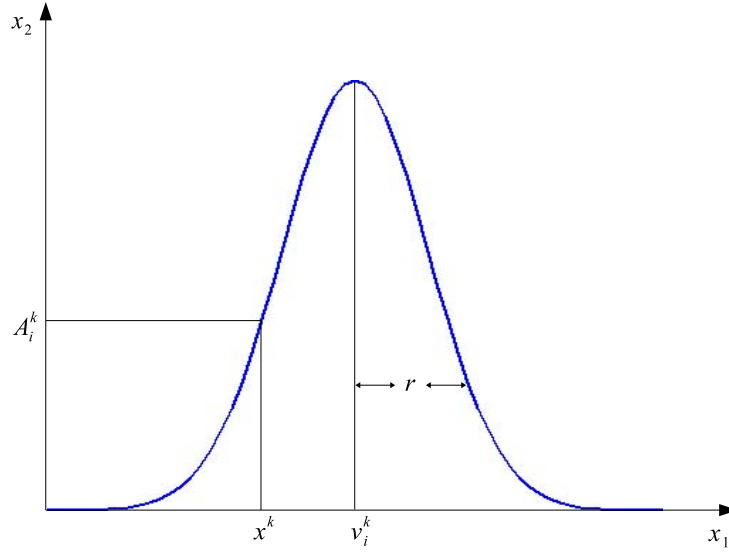


Figura 6.2: Função de pertinência Gaussiana com valor modal no centro de grupo.

6.2.2 Atualização dos conseqüentes das regras

No Aprendizado Participativo Evolutivo, a modificação ou criação de uma nova regra implica também a alteração dos parâmetros nos respectivos conseqüentes. Como a saída de cada regra representa um ajuste linear, a atualização da base de regras pode ser feita globalmente ou localmente. No caso global, em cada instante k os conseqüentes de todas as c^k regras são atualizados, e no caso local, apenas a regra mais significativa é alterada.

No modelo Takagi-Sugeno evolutivo, o conseqüente de uma regra define um ajuste representado por uma função linear do conjunto de dados, e pode ser escrito como:

$$y_i^k = \gamma_{i0} + \sum_{j=1}^p \gamma_{ij} x_j^k \quad (6.7)$$

para $i = 1, \dots, c^k$, p a dimensão do vetor de entradas e $k = 1, 2, \dots$

Os parâmetros γ_{ij} são estimados utilizando mínimos quadrados recursivo. Tomando o vetor de observações $x^k = [x_1^k, x_2^k, \dots, x_p^k] \in [0, 1]^p$ e a saída $y_i^k \in [0, 1]$ da regra

$i = 1, 2, \dots, c^k$ no instante $k = 1, 2, \dots$, a expressão (6.7) é reescrita como:

$$Y^k = X^k \gamma^k \quad (6.8)$$

sendo $\gamma^k \in \mathbb{R}^{p+1}$ o vetor dos parâmetros a ser determinado dado por:

$$(\gamma^k)^T = [\gamma_0^k \quad \gamma_1^k \quad \gamma_2^k \quad \dots \quad \gamma_p^k]$$

onde $(\cdot)^T$ denota o vetor transposto, $X^k = [1 \ x^k] \in [0, 1]^{k \times (p+1)}$ a matriz com os vetores de entrada em cada linha considerando os parâmetros livres γ_0^k , e $Y^k = [y_i^k] \in [0, 1]^k$ o vetor com as respectivas saídas.

A partir do instante $k > p + 1$, a expressão (6.8) representa um sistema linear sobredeterminado, com número de equações maior que a dimensão da variável γ^k . Dessa forma, não há garantia de solução única, e o sistema (6.8) é tal que:

$$Y^k = X^k \gamma^k + e^k \quad (6.9)$$

sendo γ^k o parâmetro do ajuste linear a ser atualizado recursivamente e e^k o respectivo erro calculado no instante k . Com isso, considerando a função objetivo:

$$J_k = J(\gamma^k) = (e^k)^T e^k = (Y^k - X^k \gamma^k)^T (Y^k - X^k \gamma^k) \quad (6.10)$$

o objetivo é determinar a solução γ^k que minimiza a função J_k .

Pela construção das matrizes X^k e Y^k , tem-se:

$$X^{k+1} = \begin{pmatrix} X^k \\ 1 \quad x^{k+1} \end{pmatrix}, \quad Y^{k+1} = \begin{pmatrix} Y^k \\ y_i^{k+1} \end{pmatrix} \quad (6.11)$$

sendo x^{k+1} a observação no instante $k + 1$ e y_i^{k+1} a respectiva saída do modelo. A solução

que minimiza o funcional J_k no instante k é escrita como (Young 1984):

$$\gamma^k = Q^k b^k \quad (6.12)$$

com $Q^k = \left[(X^k)^T X^k \right]^{-1}$, a matriz de covariância de X^k , e $b^k = (X^k)^T Y^k$.

Logo, $\gamma^{k+1} = \left[(X^{k+1})^T X^{k+1} \right]^{-1} (X^{k+1})^T Y^{k+1}$, podendo ser reescrito a partir de (6.11) como:

$$\gamma^{k+1} = \left[(X^k)^T X^k + (x^{k+1})^T x^{k+1} \right]^{-1} \left[(X^k)^T Y^k + (x^{k+1})^T y_i^{k+1} \right] = Q^{k+1} b^{k+1} \quad (6.13)$$

Comparando (6.12) e (6.13), a matriz Q^k e o vetor b^k são calculados recursivamente por:

$$(Q^{k+1})^{-1} = (Q^k)^{-1} + x^{k+1} (x^{k+1})^T \quad \text{e} \quad b^{k+1} = b^k + (x^{k+1})^T y_i^{k+1} \quad (6.14)$$

Utilizando o lema da matriz inversa (Young 1984):

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

Substituindo $A = (Q^k)^{-1}$, $C = I$, $B = x^{k+1}$, $D = (x^{k+1})^T$ e utilizando (6.14), a seguinte expressão é obtida:

$$Q^{k+1} = \left[Q^k + x^{k+1} (x^{k+1})^T \right]^{-1} = Q^k - \frac{Q^k x^{k+1} (x^{k+1})^T Q^k}{1 + (x^{k+1})^T Q^k x^{k+1}} \quad (6.15)$$

Multiplicando (6.15) por b^{k+1} pela direita e utilizando as expressões (6.12) e (6.14), tem-se a seguinte solução:

$$\gamma^{k+1} = \gamma^k + Q^{k+1} x^{k+1} \left[y_i^{k+1} - (x^{k+1})^T \gamma^k \right] \quad (6.16)$$

definindo a atualização recursiva para a solução de mínimos quadrados de J_k .

Em cada instante, a matriz de covariância Q^{k+1} é calculada por (6.15) e utilizada para obter o vetor de parâmetros γ^{k+1} em (6.16), de forma a convergir para o mínimo do erro quadrático J_k definido em (6.10) (Astrom e Wittenmark 1988).

A inicialização dos conseqüentes, isto é, a escolha dos valores iniciais de γ^0 e da matriz de covariância Q^0 , é feita de acordo com o problema. Em geral, no modelo Takagi-Sugeno evolutivo as seguintes situações são consideradas:

1. Existe um conhecimento prévio acerca do sistema. Pelo menos $p + 1$ observações são conhecidas de forma que $X^0 \in [0, 1]^{(p+1) \times (p+1)}$ e $Y^0 \in [0, 1]^{p+1}$. A inicialização, portanto, é dada por:

$$\gamma^0 = \left[(X^0)^T X^0 \right]^{-1} (X^0)^T Y^0$$

ou seja, a solução do sistema linear (6.8) para essas $p + 1$ observações. A partir daí, os parâmetros seguintes são obtidos utilizando a solução recursiva (6.16).

2. Não há conhecimento prévio algum, apenas as observações x^0 e y^0 são conhecidas. O mecanismo deve escolher os valores de γ^0 e da matriz de covariância Q^0 . A partir desses primeiros dados x^0 e y^0 , a inicialização pode ser obtida usando $\gamma^0 = [y^0 \ 0 \ \dots \ 0]$ e, portanto:

$$y^0 = [1 \ x^0] \gamma^0$$

Para a matriz de covariância, uma estratégia é inicializar Q^0 com um valor considerado alto para o sistema, isto é:

$$Q^0 = \omega I_{p+1}$$

onde I_{p+1} é a matriz identidade com ordem dada pela dimensão do vetor de parâmetros, e ω é um valor real considerado grande, como por exemplo $\omega \in [100, 10000]$ (Wellstead e Zarrop 1995).

Considerando um ajuste global nos parâmetros dos conseqüentes, em cada instante a matriz de covariância é computada para todo o sistema, e cada regra é atualizada a partir dessa matriz e da expressão para γ^k obtida em (6.16). Por outro lado, caso o ajuste seja local, para cada regra uma matriz de covariância é computada, e a atualização é feita apenas na regra cujo antecedente foi modificado no instante k .

Como será visto na próxima seção, no caso onde há alteração no número de grupos os parâmetros do conseqüente da nova regra podem ser inicializados a partir das regras que permaneceram inalteradas. A idéia nesse caso é aproveitar os centros de grupo e os ajustes existentes para estimar os parâmetros da nova regra (Angelov e Filev 2004).

6.3 Algoritmo de Aprendizado Participativo Evolutivo

O Aprendizado Participativo Evolutivo baseia-se na obtenção recursiva dos centros de grupo e no cálculo dos antecedentes e conseqüentes. À cada instante k no qual uma observação x^k é fornecida, o algoritmo do APE pode ser dividido em duas etapas básicas:

1. Atualização dos antecedentes das regras, determinando os centros dos grupos e as respectivas funções de pertinência;
2. Atualização dos conseqüentes das regras utilizando mínimos quadrados recursivo.

Apenas a nova observação e as regras atuais são utilizadas na atualização do sistema. Dessa forma, o APE modifica os centros nos antecedentes a partir da combinação linear definida no aprendizado participativo, e modifica os conseqüentes a partir dos mínimos quadrados recursivo.

A inicialização das regras pode ser feita tanto a partir do agrupamento de um conjunto de treinamento como a partir de, pelo menos, duas observações. Caso utilize o

conjunto de treinamento para a inicialização, o primeiro ajuste das regras é obtido após o processo de agrupamento desse conjunto de treinamento, e o algoritmo APE segue de forma recursiva para as observações seguintes. Caso seja inicializado a partir das primeiras observações, dois pontos são escolhidos como centros iniciais e o primeiro ajuste pode ser obtido a partir da saída conhecida dessas observações, inicializando os consequentes como mostrado na Seção 6.2.2.

No algoritmo de agrupamento não supervisionado o número de grupos pode mudar e, portanto, deve-se definir também o comportamento do algoritmo caso o número de regras seja modificado, ou seja, definir quais serão os centros e os respectivos parâmetros do consequente caso um novo grupo seja criado. Nesse caso, (Angelov e Filev 2004) sugere que os parâmetros do consequente sejam obtidos a partir das outras regras existentes, considerando os graus de ativação μ_i^k e os parâmetros γ_i^k , $i = 1, 2, \dots, c^k$. Assim, o ajuste da nova regra $c^k + 1$ é obtido por:

$$\gamma_{c^k+1}^k = \sum_{j=1}^{c^k} \mu_j^k \gamma_j^k \quad (6.17)$$

sendo o centro representado pela observação que gerou a criação do novo grupo, isto é:

$$v_{c^k+1}^k = x^k \quad (6.18)$$

Uma outra possibilidade seria inicializar a nova regra usando a estratégia comentada na Seção 6.2.2, onde a matriz de covariância é obtida a partir de um valor arbitrariamente alto.

Analogamente, no caso onde algum grupo é excluído, isto é, quando existem dois centros de grupo distintos v_i^k e v_j^k com compatibilidade entre eles considerada alta, $\rho_{ij}^k \geq \lambda$, os parâmetros do grupo resultante podem ser obtidos a partir desses grupos

redundantes i e j da seguinte forma:

$$\gamma_r^k = \mu_i^k \gamma_i^k + \mu_j^k \gamma_j^k \quad (6.19)$$

sendo que o respectivo centro v_r^k pode ser obtido por uma expressão análoga à utilizada na atualização dos centros, ou seja:

$$v_r^k = v_i^k + \alpha \rho_{ij}^k (v_j^k - v_i^k) \quad (6.20)$$

Portanto, um centro redundante é excluído contribuindo na atualização do outro. Um caso particular de (6.20) seria calcular o centro v_r^k do grupo resultante como o valor médio dos centros dos grupos redundantes:

$$v_r^k = \frac{1}{2} v_i^k + \frac{1}{2} v_j^k$$

ou, simplesmente, manter um dos grupos e excluir o outro redundante.

Em relação aos parâmetros $\tau \in [0, 1]$, $\lambda \in [0, 1]$ e a sensibilidade do índice de alerta $\beta \in [0, 1]$, os valores podem ser escolhidos satisfazendo a análise do agrupamento participativo feita no Capítulo 4, tomando a desigualdade (4.7) dada por:

$$\tau \leq \beta(1 - \lambda) \quad (6.21)$$

e uma sugestão seria escolher τ e considerar $\beta = \tau$ e $\lambda = \frac{1 - \tau}{2}$. Relacionar os parâmetros é uma maneira de reduzir o número de escolhas a serem determinadas no APE, garantindo as propriedades do agrupamento participativo analisadas no Capítulo 4.

A Figura 6.3 mostra o diagrama que resume os passos principais do algoritmo APE, que é mostrado logo em seguida.

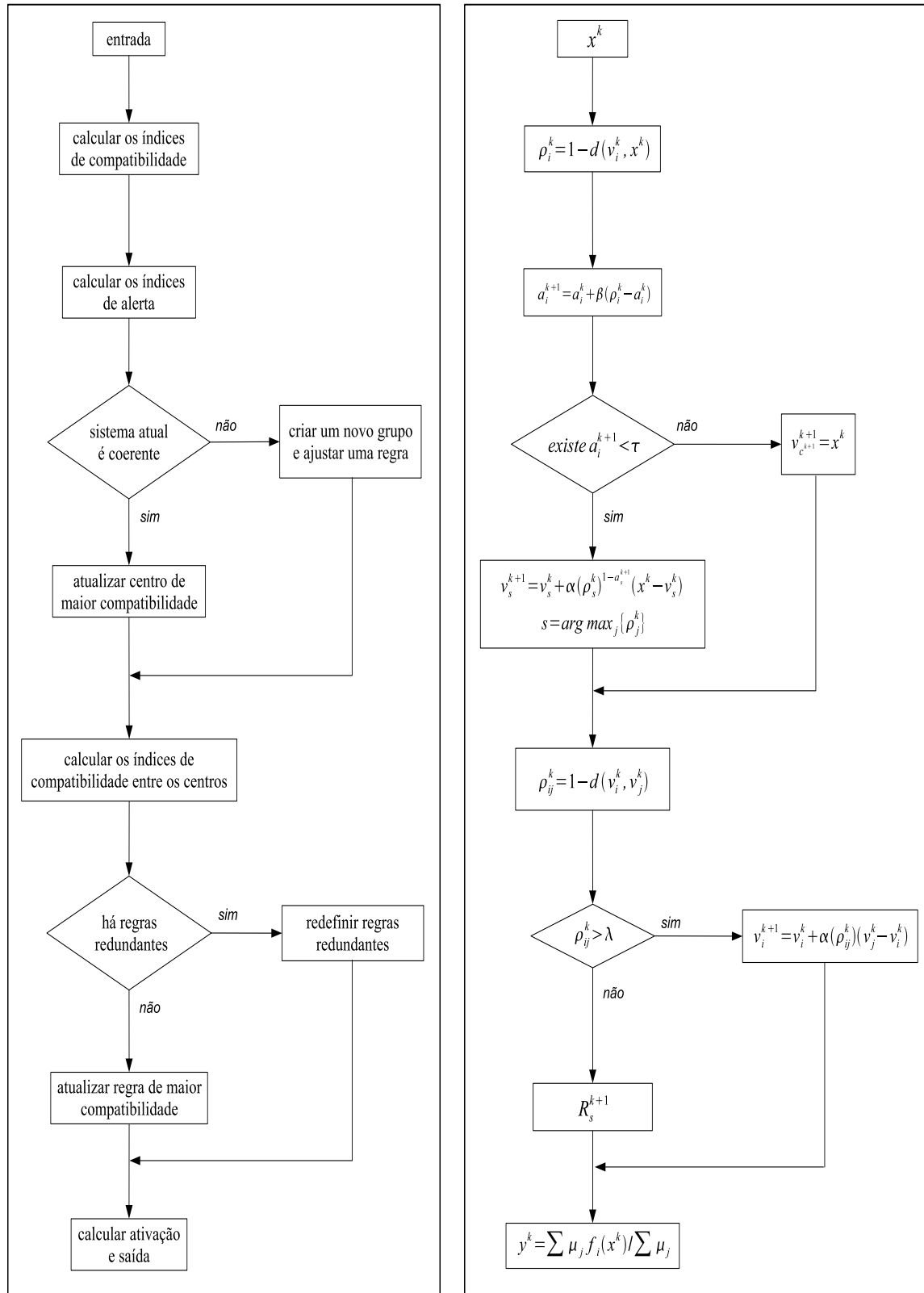


Figura 6.3: Diagrama com os passos do APE.

ALGORITMO APRENDIZADO PARTICIPATIVO EVOLUTIVO

Inicializar o número de grupos $c^1 \geq 2$ e os centros $v_i^1 \in [0, 1]^p$

Definir o índice de alerta $a_i^1 = 0, i = 1, 2, \dots, c^1$.

Escolher os parâmetros $\alpha \in [0, 1], \beta \in [0, 1], \tau \in [0, 1], \lambda \in [0, 1]$ e $r \in [0, 1]$.

Para $x^k \in [0, 1]^p, k = 1, 2, \dots$

Para $i = 1, 2, \dots, c^k$

Calcular a compatibilidade: $\rho_i^k = 1 - d(v_i^k, x^k)$;

Calcular o alerta: $a_i^{k+1} = a_i^k + \beta(1 - \rho_i^k - a_i^k)$;

Calcular o grau de ativação: $\mu_i^k = e^{-r\|v_i^k - x^k\|^2}$;

Fim Para

Se $a_i^{k+1} \geq \tau, \forall i = 1, 2, \dots, c^k$ **então**

Acrescentar um grupo: $c^k = c^k + 1$;

Inicializar os parâmetros do conseqüente: $\gamma_{c^k}^k = \sum_{j=1}^{c^k-1} \mu_j^k \gamma_j^k$

x^k é centro do novo grupo: $v_{c^k}^k = x^k$;

Senão

Encontrar o grupo mais compatível: $s = \arg \max_{j=1, \dots, c^k} \{\rho_j^k\}$;

Atualizar o grupo s : $v_s^{k+1} = v_s^k + \alpha(\rho_s^k)^{1-a_s^{k+1}}(x^k - v_s^k)$;

Atualizar os parâmetros do conseqüente: $\gamma_s^k = \gamma_s^{k-1} + Q^k x^k \left[y_s^k - (x^k)^T \gamma_s^{k-1} \right]$;

Fim Se

Para $i = 1, 2, \dots, (c^k - 1)$ e $j = (i + 1), \dots, c^k$

Calcular compatibilidade entre centros v_i e v_j : $\rho_{ij}^k = 1 - d(v_i^k, v_j^k)$;

Se $\rho_{ij}^k \geq \lambda$ para algum j **então**

Redefinir v_i^k usando os grupos i e j e excluir v_j^k ;

Excluir um grupo: $c^k = c^k - 1$;

Fim Se

Fim Para

Calcular saída do modelo: $y^k = \frac{\sum_{i=1}^{c^k} \mu_i^k \gamma_i^k x^k}{\sum_{i=1}^{c^k} \mu_i^k}$

No algoritmo, a redefinição dos grupos quando uma regra é excluída pode ser feita a partir das estratégias comentadas nessa seção. No caso mais simples, um dos grupos redundantes é excluído e o outro é mantido, não alterando a inclinação do respectivo ajuste no conseqüente. Essa é a estratégia utilizada na obtenção dos resultados presentes no Capítulo 7.

6.4 Resumo

Este capítulo apresentou a proposta desse trabalho, intitulada Aprendizado Participativo Evolutivo. Na proposta, o agrupamento participativo é utilizado para a atualização dos centros de grupo e o modelo Takagi-Sugeno evolutivo é utilizado para a construção de um modelo fuzzy.

A principal diferença entre os algoritmos APE e eTS está na forma como o agrupamento é realizado. Diferentemente do eTS, o APE usa o aprendizado participativo na adaptação dos grupos e, conseqüentemente, acrescenta uma nova interpretação ao sistema em estudo.

No próximo capítulo o algoritmo proposto será aplicado para previsão de séries temporais. O desempenho do APE será comparado com o algoritmo eTS e com modelos presentes na literatura.

Capítulo 7

Resultados

7.1 Introdução

Nesse capítulo, o desempenho do algoritmo APE é verificado em problemas de previsão de séries temporais. No problema de previsão, as realizações em instantes anteriores são conhecidas e o modelo deve estimar as observações nos instantes seguintes, utilizando o conjunto de dados e os parâmetros atualizados do modelo.

A partir das observações anteriores x^{k-j} , o objetivo é estimar os valores das observações seguintes, onde o valor de $j \in [0, k-1]$ representa número de atrasos utilizados para a previsão da série no instante $k+1$. Uma das formas de determinar o número de atrasos, ou seja, a escolha do valor de j , é a partir da função de autocorrelação parcial, onde busca-se detectar os instantes anteriores mais correlacionados com os instantes a serem previstos. Dessa forma, em um modelo fuzzy funcional Takagi-Sugeno, os ajustes dos parâmetros do conseqüente de cada regra são baseados nos valores das variáveis $x^k, x^{k-1}, \dots, x^{k-j}$, em instantes anteriores.

Por outro lado, o número de instantes seguintes que serão previstos também pode variar. Nesse caso, a partir de $x^k, x^{k-1}, \dots, x^{k-j}$, o objetivo é estimar os valores de $x^{k+1}, x^{k+2}, \dots, x^{k+m}$ sendo $m = 1, 2, \dots$ o número de passos à frente na previsão. Se $m = 1$, o problema é conhecido como previsão um passo à frente, se $m > 1$ o problema é conhecido como previsão m passos à frente (Nelles 2000). Vale notar que, em problemas de previsão vários passos à frente, se $m > j$ existe um instante no qual o modelo utiliza os próprios

valores previstos para a estimativa do instante seguinte, ocorrendo o que é chamado de previsão sobre previsão.

Nesse trabalho, o desempenho do APE é verificado em problemas de previsão um passo à frente, ou seja, $m = 1$. A escolha do número j de instantes anteriores considerados varia de acordo com o problema. Os resultados são analisados em dois casos: a conhecida série Box-Jenkins e uma série horária de carga elétrica de uma empresa da região sudeste do Brasil.

Como medida de desempenho, os modelos são comparados utilizando a raiz quadrada do erro quadrático médio (REQM) e o índice de erro não dimensional (IEND), esse último calculado como a razão entre a REQM e o desvio padrão dos valores que deseja-se prever, como mostrado nas expressões (7.1) e (7.2):

$$\text{REQM} = \left(\frac{1}{N} \sum_{k=1}^N (\bar{y}^k - y^k)^2 \right)^{\frac{1}{2}} \quad (7.1)$$

$$\text{IEND} = \frac{\text{REQM}}{\text{std}(\bar{y}^k)} \quad (7.2)$$

sendo que N é o número de instantes previstos, y^k é a saída do modelo de previsão, $\bar{y}^k$ é a saída real esperada e $\text{std}(\cdot)$ é o desvio padrão.

As análises do agrupamento participativo feitas no Capítulo 4 são usadas na escolha dos parâmetros τ e λ utilizados na implementação do APE. Como ilustrado na Figura 7.1, tomando o limiar do índice de alerta τ , o limiar λ da compatibilidade entre os centros é definido como $\lambda = \frac{1 - \tau}{2}$. A dispersão r da função Gaussiana ajustada nos antecedentes varia em cada problema.

O desempenho do algoritmo APE proposto nesse trabalho é comparado com o algoritmo eTS e suas extensões conhecidas como simpl_eTS e xTS. Além desses modelos, o desempenho do APE é comparado também com o modelo de rede neural multicamada

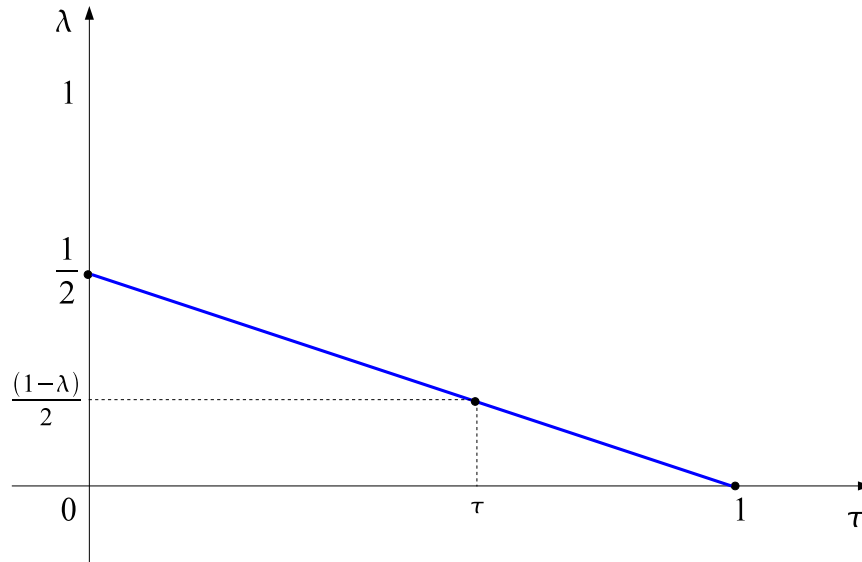


Figura 7.1: Escolha do parâmetro λ .

com retropropagação de erro MLP (Rumelhart et al. 1986) e com o sistema de inferência neurofuzzy ANFIS (Jang 1993). Esses algoritmos vêm sendo bastante utilizados em problemas de previsão de séries temporais.

7.2 Previsão da Série Box-Jenkins

A série Box-Jenkins é bastante conhecida na literatura sendo utilizada na comparação de modelos de agrupamento e previsão. A série consiste em 290 valores de entradas e saídas resultantes de um experimento em forno a gás. A entrada x^k consiste no fluxo de gás metano e a saída y^k representa a respectiva concentração de gás carbônico. A equação que melhor descreve esse experimento é dada por (Box, Jenkins e Reinsel 1994):

$$y^k = f(y^{k-1}, x^{k-4})$$

Os resultados mostrados na Tabela 7.2 correspondem às simulações utilizando os 200 primeiros valores da série como conjunto de treinamento para a previsão dos últimos

90 valores. O APE é comparado ao desempenho do algoritmo eTS e duas de suas variações: o xTS (Angelov e Zhou 2006), que utiliza o eTS com o raio adaptativo, e o simpl_eTS (Angelov e Filev 2005), que ao invés da função Gaussiana utiliza a função de Cauchy como função de pertinência do antecedente. Os resultados da rede neural multicamada MLP (Rumelhart et al. 1986) e do sistema de inferência neurofuzzy ANFIS (Jang 1993) foram obtidos a partir do toolbox de redes neurais do Matlab.

Os melhores resultados obtidos com a rede MLP foram utilizando 5 neurônios na camada intermediária, treinados a partir do algoritmo de retropropagação do erro. Os valores dos pesos foram inicializados aleatoriamente, a taxa de aprendizagem foi inicializada como $\eta = 0.01$ e o número máximo de iterações considerado foi igual a 10000. No sistema ANFIS, os melhores resultados foram considerando $\eta_i = 0.01$ o tamanho inicial do passo, $si = 1.1$ a taxa de crescimento, $sd = 0.9$ a taxa de decrescimento e 1000 o número máximo de iterações. O resultado do xTS foi obtido a partir de uma implementação disponibilizada pelo desenvolvedor do algoritmo, e os resultados dos algoritmo eTS e simpl_eTS foram extraídos de um mesmo experimento divulgado em (Angelov e Filev 2005).

Tabela 7.1: Comparação na previsão da série Box-Jenkins.

Modelo	Número de regras (ou nós)	<i>REQM</i>	<i>IEND</i>
APE	4	0.0191	0.1189
ANFIS (Jang 1993)	7	0.0207	0.1294
MLP (Rumelhart et al. 1986)	5	0.0211	0.1319
xTS (Angelov e Zhou 2006)	6	0.0375	0.2397
simpl_eTS (Angelov e Filev 2005)	3	0.0485	0.3023
eTS (Angelov e Filev 2005)	5	0.0490	0.3057

O APE detectou 3 grupos no conjunto de treinamento e durante o período de previsão mais uma regra foi adicionada. A Figura 7.2 mostra o comportamento do índice de alerta, onde o APE conseguiu o melhor resultado considerando o parâmetro $\tau = 0.16$.

A Figura 7.3 mostra o gráfico da saída do APE utilizando os parâmetros $\tau =$

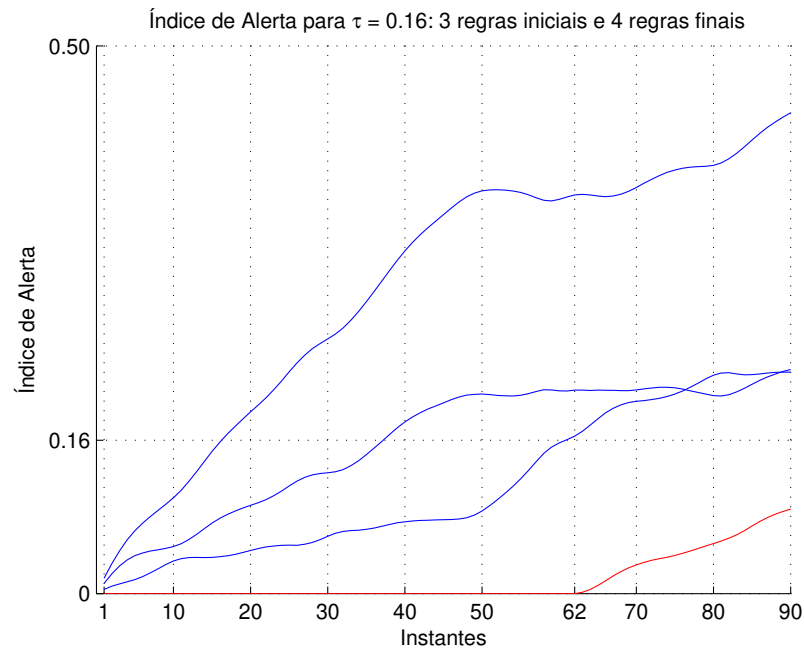


Figura 7.2: Comportamento do índice de alerta na previsão da série Box-Jenkins.

0.16 , $r = 0.4$, $\lambda = \frac{1 - \tau}{2} = 0.42$ e $\beta = \tau = 0.16$. Ainda pelas Figuras 7.2 e 7.3, pode-se observar também que um novo grupo foi adicionado no instante 62 da previsão, totalizando 4 regras ajustadas.

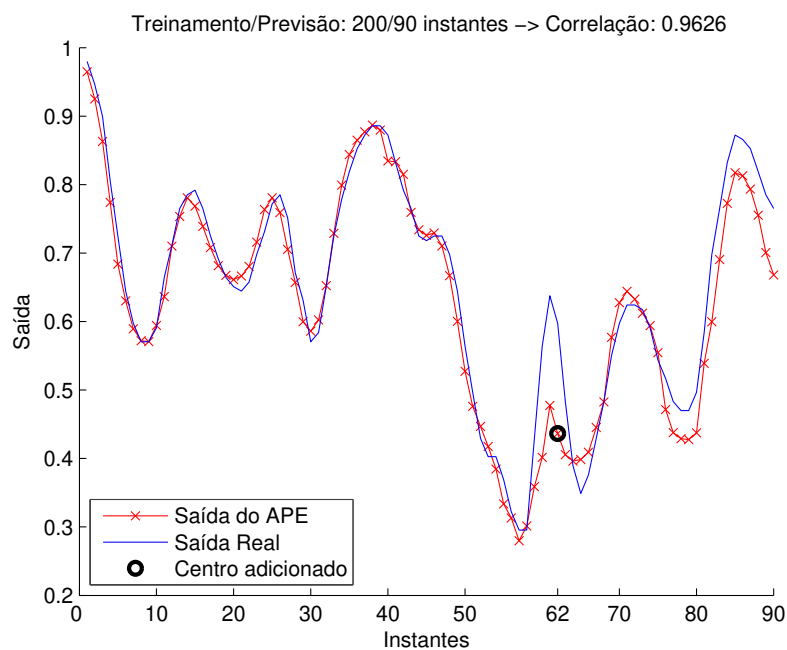


Figura 7.3: Saída do APE na previsão da série Box-Jenkins.

Vale destacar o número reduzido de regras que o APE ajusta para obter os melhores resultados. A partir do algoritmo definido no Capítulo 6, pode-se notar que diminuir o valor de τ possibilita a criação de um número maior de regras. No entanto, aumentar o número de regras no APE não necessariamente diminui o valor da REQM.

7.3 Previsão de Carga Elétrica

A previsão de carga é de suma importância para um bom planejamento de energia elétrica. A programação da geração e das intervenções no sistema, bem como a execução da operação em tempo real são processos cuja qualidade pode ser consideravelmente aumentada a partir de um bom modelo de previsão dos valores de carga esperados (Lima et al. 2007).

Nessa seção, as comparações com o APE são feitas na previsão de carga a curto prazo de uma empresa de energia elétrica. Os dados de carga usados são horários e expressos em kilowatts (kW), correspondentes aos 31 dias do mês de Agosto de 2000.

Para a previsão da carga uma hora à frente foram utilizadas 2 horas anteriores, como consequência da função de autocorrelação parcial mostrada na Figura 7.4. Considerando essa função calculada para os 36 primeiros valores da série, o gráfico sugere os dois primeiros passos anteriores como os mais correlacionados, considerando a linha pontilhada do erro padrão.

Os modelos APE, eTS, xTS, MLP e ANFIS são avaliados para a previsão do último dia do mês de Agosto de 2000, considerando os 30 primeiros dias como período de treinamento. Dessa forma, na série com 744 valores horários, as primeiras 720 horas servem de treinamento para a previsão dos últimos 24 valores.

Durante o processo de treinamento e previsão, os dados de entrada foram

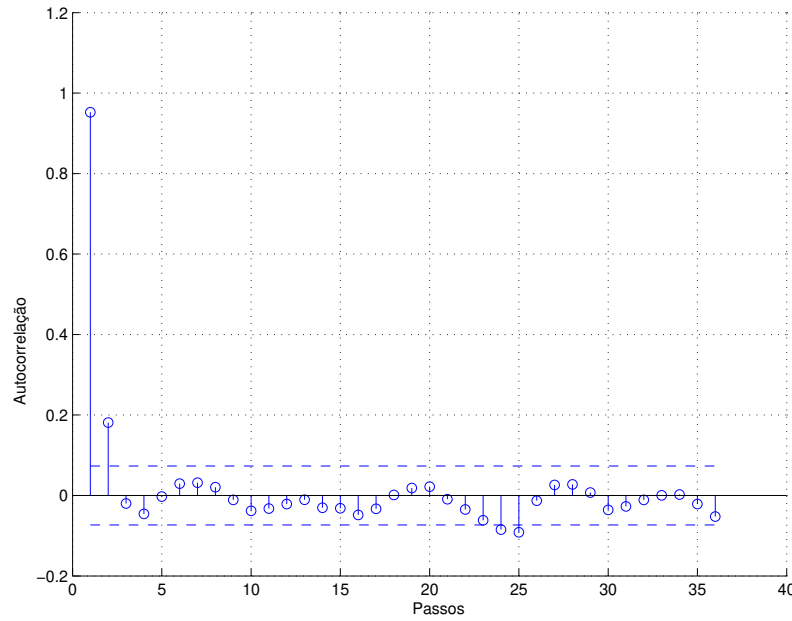


Figura 7.4: Função de autocorrelação parcial da série horária de carga elétrica.

normalizados no intervalo $[0, 1]$ da seguinte forma:

$$\bar{x}^k = \frac{x^k - \min(x^k)}{\max(x^k) - \min(x^k)}$$

onde $\min(x^k)$ denota o menor valor da série x^k e $\max(x^k)$ o maior valor dessa série, computados a partir de todo o histórico.

Os parâmetros utilizados no APE foram $\tau = 0.30$, $r = 0.25$ e $\lambda = 0.35$. Os resultados do eTS e do xTS foram obtidos a partir de uma implementação disponibilizada pelo desenvolvedor do algoritmo (Angelov e Zhou 2006). A rede MLP e o sistema ANFIS foram executados no Matlab conforme descrito na seção anterior.

A Tabela 7.2 mostra o APE como um algoritmo bastante competitivo, onde se destaca novamente o reduzido número de regras que são ajustadas durante o processo. Vale lembrar que o APE é não supervisionado tanto no período de treinamento quanto no período de previsão, ou seja, o número inicial de grupos utilizados no processo de previsão é decidido pelo agrupamento participativo, executado para o período de treinamento. Assim, todo o processo de treinamento e previsão é não supervisionado.

Tabela 7.2: Resultados na previsão da série horária de carga elétrica.

Modelo	Número de regras (ou nós)	<i>REQM</i>	<i>IEND</i>
APE	5	0.0508	0.1888
eTS (Angelov e Filev 2005)	6	0.0517	0.1922
ANFIS (Jang 1993)	9	0.0541	0.2012
MLP (Rumelhart et al. 1986)	5	0.0552	0.2055
xTS (Angelov e Zhou 2006)	4	0.0562	0.2089

Como pode-se notar na Figura 7.5, que mostra o comportamento dos índices de alerta para os 24 valores previstos de carga elétrica, o APE obteve três grupos durante o período de treinamento, e durante o período de previsão duas outras regras foram adicionadas nos instantes 9 e 14, totalizando 5 regras ajustadas. Os instantes no qual novas

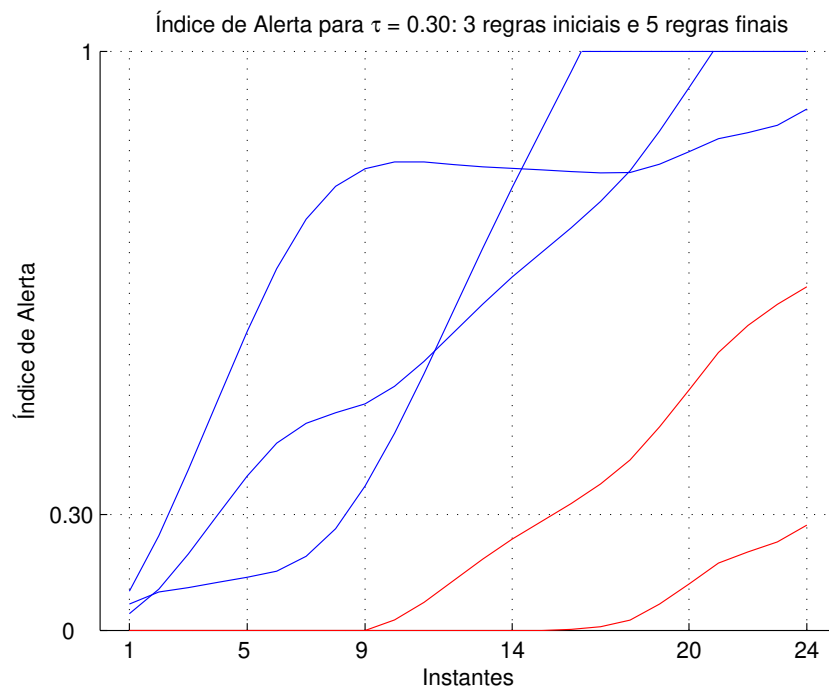


Figura 7.5: Comportamento do índice de alerta na série horária de carga elétrica.

regras foram adicionadas também estão mostrados na Figura 7.6, que mostra o gráfico com a saída do APE e a saída real. A correlação entre a série real a série prevista pelo APE ficou por volta de 98%.

Nas Figuras 7.7 e 7.8 são mostradas as saídas dos algoritmos eTS e xTS res-

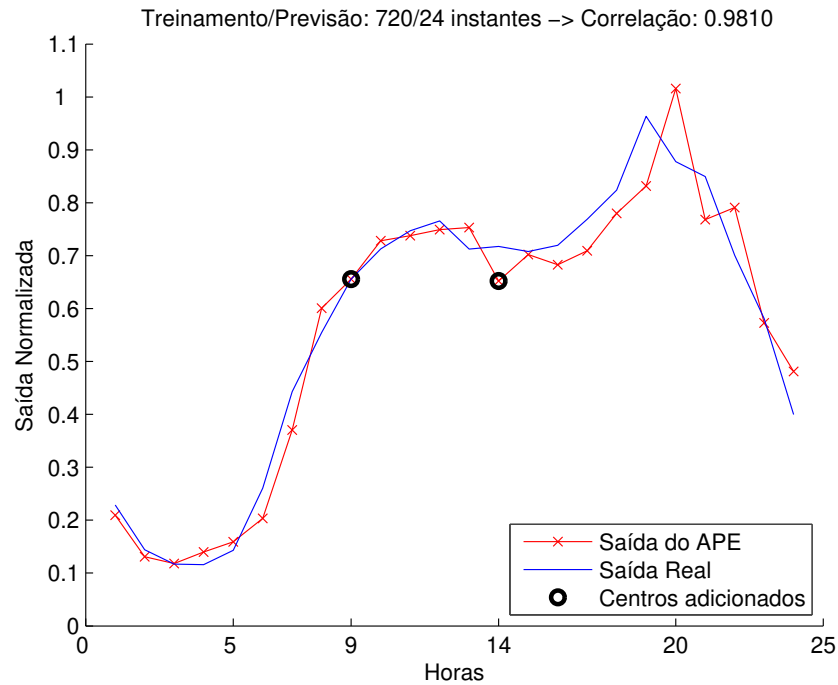


Figura 7.6: Saída do APE na previsão da série horária de carga elétrica.

pectivamente, sendo que na Figura 7.9 é mostrada a saída da MLP para o problema e na Figura 7.10 é mostrada a previsão obtida com o sistema ANFIS. No período diário da carga elétrica prevista nota-se, ainda pelas Figuras 7.7-7.10, que a maior dificuldade dos modelos foi no ponto de máximo da série em questão. Nesse ponto, é importante que o algoritmo seja flexível e detecte também as maiores variações da série. As saídas mostradas estão normalizadas.

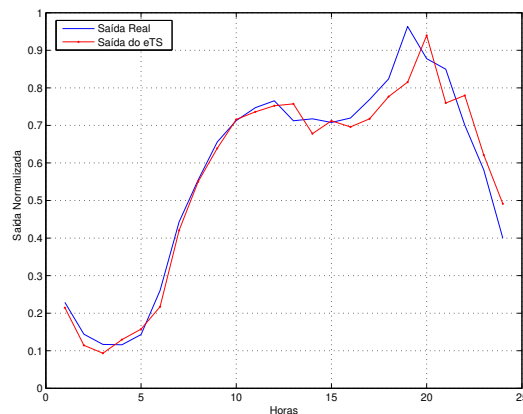


Figura 7.7: Saída do eTS na previsão da série horária de carga elétrica.

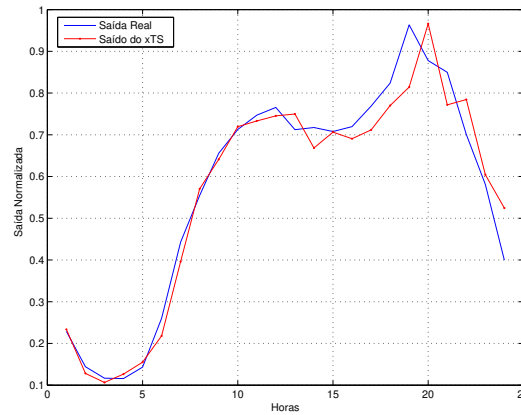


Figura 7.8: Saída do xTS na previsão da série horária de carga elétrica.

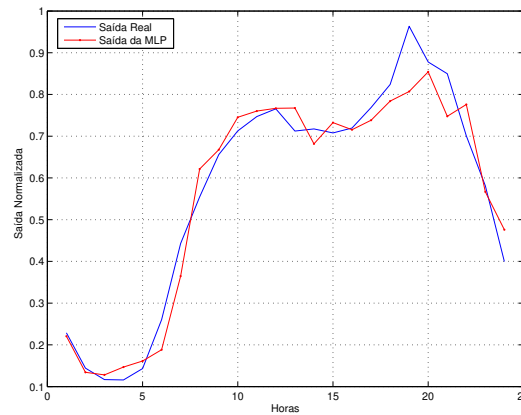


Figura 7.9: Saída da MLP na previsão da série horária de carga elétrica.

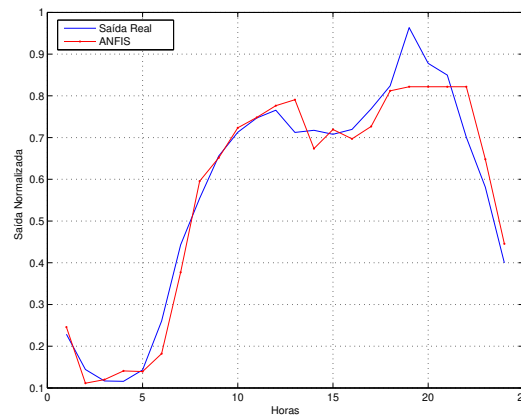


Figura 7.10: Saída do ANFIS na previsão da série horária de carga elétrica.

Além dos parâmetros τ , r e λ , o número inicial de grupos também representa um parâmetro de influência no modelo. A partir do algoritmo APE definido no Capítulo 6, o número inicial de grupos pode ser obtido por diversas estratégias diferentes. Algoritmos

como o Fuzzy C-Means (Bezdek 1981) podem ser utilizados para fixar um número inicial de grupos em casos onde um período anterior à previsão é conhecido.

Nesse trabalho, o número inicial de grupos do APE foi obtido de forma não supervisionada a partir do agrupamento participativo definido no Capítulo 4. Conhecido um período anterior ao que será previsto, o algoritmo de agrupamento fornece também o ajuste inicial da base de regras, que é obtido a partir do algoritmo de mínimos quadrados aplicado aos dados que atualizaram os grupos no agrupamento do período de treinamento.

Os resultados na Seção 7.3 mostram o melhor valor obtido pelo APE, no caso onde era conhecido 1 mês de valores da série. Nesse sentido, é interessante verificar os resultados do APE utilizando também outros períodos de treinamento.

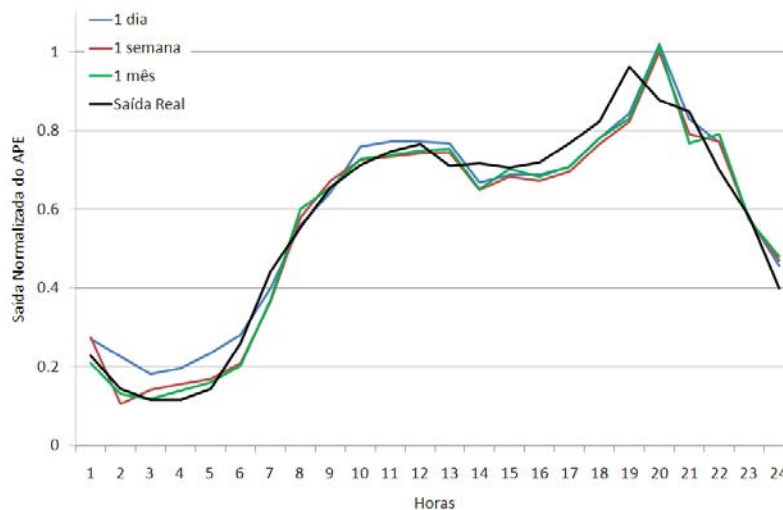


Figura 7.11: Saídas para diversos períodos de treinamento

Tabela 7.3: Comparação do APE utilizando diversos períodos para o treinamento

Treinamento	<i>REQM</i>	<i>IEND</i>
1 dia	0.0542	0.2018
1 semana	0.0586	0.2180
1 mês	0.0508	0.1888

7.4 Resumo

Este capítulo mostrou alguns resultados da implementação do algoritmo APE aplicada em problemas de previsão de séries temporais. Os resultados nas previsões da série Box-Jenkins e da série de carga mostraram o APE como um algoritmo bastante competitivo, ajustando um número reduzido de regras.

Os parâmetros da etapa do agrupamento participativo foram obtidos a partir de testes computacionais. Os valores de τ , λ e β variaram de acordo com o comportamento da série prevista. O algoritmo APE mostrou-se estável, e capaz de adaptar-se de maneira bastante eficiente independentemente do tamanho do conjunto de treinamento.

Capítulo 8

Conclusão

As abordagens evolutivas são bastante eficientes em problemas de previsão em tempo real, onde os dados são fornecidos de forma gradativa e o algoritmo deve trabalhar de maneira adaptativa. Em cada instante, tanto a estrutura quanto os parâmetros que descrevem o algoritmo são atualizados em tempo real.

Nesse sentido, esse trabalho propôs um modelo fuzzy funcional evolutivo baseado na noção do aprendizado participativo. Chamado Aprendizado Participativo Evolutivo, esse modelo é inspirado no modelo Takagi-Sugeno evolutivo conhecido como eTS, e inclui um mecanismo de agrupamento não supervisionado capaz de identificar dinamicamente o número de grupos, juntamente com um mecanismo que atualiza em tempo real a estrutura de regras.

Uma análise do aprendizado participativo foi feita com o intuito de obter uma versão dinâmica do algoritmo de agrupamento participativo. Aplicando esse algoritmo de agrupamento na construção de uma base de regras, um eficiente algoritmo é obtido unindo o conceito de aprendizagem para atualização dos antecedentes com o conceito de mínimos quadrados para a atualização recursiva dos conseqüentes. O Aprendizado Participativo Evolutivo proposto nesse trabalho é um algoritmo não supervisionado, que atualiza o número de grupos e os respectivos parâmetros de acordo com a entrada dos dados.

As análises do agrupamento participativo, ainda não encontradas na literatura, são também uma importante contribuição desse trabalho. Os resultados estão presentes no Capítulo 3 desse trabalho sobre a convergência do aprendizado participativo quando visto

como um modelo iterativo.

O APE mostrou-se bastante competitivo e capaz de obter bons resultados de forma parcimoniosa, com um número reduzido de regras, quando aplicado em problemas de previsão em tempo real de séries temporais.

No problema de previsão, o desafio está em obter uma estimativa precisa para a previsão de curto prazo em tempo real, e utilizar as estimativas obtidas anteriormente na atualização do algoritmo em instantes seguintes. Nesse sentido, o desempenho na previsão da série Box-Jenkins e na previsão da série de carga elétrica mostraram o APE como um modelo leve e eficiente, onde as operações envolvidas podem ser facilmente computadas. Portanto, tem-se o APE como um poderoso algoritmo a ser aplicado em problemas reais. Modelos fuzzy funcionais evolutivos vêm sendo mostrados com êxito na literatura recente. Nesse aspecto, o APE proposto nesse trabalho é visto como uma promissora alternativa, a continuar sendo desenvolvida futuramente.

Seguindo a proposta inicial para o APE presente nesse trabalho, alguns resultados acerca da estabilidade do algoritmo ainda podem ser discutidos. Devido à relação mostrada entre os parâmetros do algoritmo de agrupamento utilizados no APE, um dos objetivos futuros é desenvolver mais precisamente a relação entre o número de grupos e as escolhas iniciais dos parâmetros do agrupamento participativo.

Assim como na extensão do eTS conhecida como xTS, o APE também pode ser expandido considerando o raio como um parâmetro adaptativo. Dessa forma, o cálculo dos antecedentes das regras dependeria também dos valores dos respectivos índices de alerta, sendo interessante analisar melhor esse conceito e considerá-lo como uma proposta futura para a expansão do Aprendizado Participativo Evolutivo.

Referências Bibliográficas

- Magalhães, M. (2004). *Redes neurais, metodologias de agrupamento e combinação de previsores aplicados à previsão de vazões naturais*, Dissertação de Mestrado, FEEC–Unicamp, Campinas, Brasil.
- Angelov, P. (2002). *Evolving Rule-based Models: A Tool for Design of Flexible Adaptive Systems*, 1ª edição, Springer-Verlag, NY, EUA.
- Angelov, P. e Filev, D. (2004). An approach to online identification of Takagi-Sugeno fuzzy models, *IEEE Transactions on Systems, Man and Cybernetics, Part B* **34**(01): 484–498.
- Angelov, P. e Filev, D. (2005). Simpl_eTS: A simplified method for learning evolving Takagi-Sugeno fuzzy models, *Proceedings of the 14th IEEE International Conference on Fuzzy Systems FUZZ-IEEE*, Reno, EUA, pp. 1068–1073.
- Angelov, P., Filev, D. e Kasabov, N. (Eds) (2008). *Evolving Intelligent Systems: Methodology and Applications*, Wiley-Interscience/IEEE Press.
- Angelov, P. e Kasabov, N. (2006). Evolving intelligent systems – eIS, *IEEE Systems, Man and Cybernetics eNewsletter* (15): 1–13.
- Angelov, P., Xydeas, C. e Filev, D. (2004). On-line identification of MIMO evolving Takagi-Sugeno fuzzy models, *Proceedings of the International Joint Conference on Neural Networks and International Conference on Fuzzy Systems, IJCNN-FUZZ-IEEE*, Budapest, Hungary, pp. 55–60.
- Angelov, P. e Zhou, X. (2006). Evolving fuzzy systems from data streams in real-time, *Proceedings of the International Symposium on Evolving Fuzzy Systems – EFS'06*, IEEE Press, Ambleside, Lake District, UK, pp. 29–35.
- Astrom, K. e Wittenmark, B. (1988). *Adaptive Systems*, 1ª edição, Addison-Wesley, EUA.

- Bazaraa, M., Sherali, H. e Shetty, C. (2006). *Nonlinear Programming: Theory and Algorithms*, 3^a edição, Wiley-Interscience.
- Bezdek, J. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers, Norwell, MA, EUA.
- Bezdek, J. e Pal, S. (1992). *Fuzzy Models for Pattern Recognition: Methods That Search for Structures in Data*, 1^a edição, IEEE Press, Piscataway, NY, EUA.
- Bezdek, J., Tsao, E. e Pal, N. (1992). Fuzzy Kohonen clustering networks, *IEEE International Conference on Fuzzy Systems* **8**(12): 1035–1043.
- Box, G., Jenkins, G. e Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control*, 3^a edição, Prentice Hall.
- Brouwer, R. (2000). A fuzzy recurrent artificial neural network (FRANN) for pattern classification, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **08**(05): 525–538.
- Buragohain, M. e Mahanta, C. (2008). A novel approach for ANFIS modelling based on full factorial design, *Applied Soft Computing* **08**(01): 609–625.
- Chen, D.-W. e Zhang, J.-P. (2005). Time series prediction based on ensemble ANFIS, *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*, Vol. 06, Beijing, China, pp. 3552–3556.
- Chiu, S. (1994). Fuzzy model identification based on cluster estimation, *Journal of Intelligence and Fuzzy Systems: Applications in Engineering and Technology* **02**(03): 267–278.
- Conte, S. e de Boor, C. (1981). *Elementary Numerical Analysis - An Algorithmic Approach*, 3^a edição, McGraw-Hill, NY, EUA.
- Duda, R. e Hart, P. (1973). *Pattern Classification and Scene Analysis*, 1^a edição, Wiley Interscience, NY, EUA.
- Dvoretzky, A. (1954). On stochastic approximation, *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, pp. 39–55.
- Er, M. e Wu, S. (2002). A fast learning algorithm for parsimonious fuzzy neural systems, *Fuzzy Sets and Systems* **126**(3): 337–351.

- Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, 3^a edição, John Wiley and Sons.
- Figueiredo, M., Ballini, R., Soares, S., Andrade, M. e Gomide, F. (2004). Learning algorithms for a class of neurofuzzy network and application, *IEEE Transactions on Systems, Man and Cybernetics–Part C* **34**(3): 293–301.
- Figueiredo, M. e Gomide, F. (1998). Fuzzy system design using neurofuzzy networks, *Proceedings of the 7th Information Processing and Management of Uncertainty on Knowledge-Based Systems* **02**: 1416–1422.
- Filev, D. e Angelov, P. (1992). Fuzzy optimal control, *Fuzzy Sets and Systems* **47**(2): 151–156.
- Gath, I. e Geva, A. (1989). Unsupervised optimal fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(07): 773–781.
- Geva, A. (1999). Non-stationary time series prediction using fuzzy clustering, *18th International Conference of the North American Fuzzy Information Processing Society - NAFIPS*, NY, EUA, pp. 413–417.
- Geva, A., Steinberg, Y., Bruckmair, S. e Nahum, G. (2000). A comparison of cluster validity criteria for a mixture of normal distributed data, *Pattern Recognition Letters* **21**(6-7): 511–529.
- Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, 1^a edição, Addison-Wesley Publishing Company, MA, EUA.
- Haykin, S. (1998). *Neural Networks – A Comprehensive Foundation*, 2^a edição, Prentice Hall, NY, EUA.
- Hertz, J., Krogh, A. e Palmer, R. (1991). *Introduction to the Theory of Neural Computation*, Santa Fe Institute Studies in the Sciences of Complexity, 1^a edição, Addison-Wesley Publishing, Redwood City, CA.
- Holland, J. (1975). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*, 1^a edição, MIT Press, Massachusetts, EUA.
- Jang, J. (1993). ANFIS: Adaptive network-based fuzzy inference systems, *IEEE Transactions on Systems, Man and Cybernetics* **23**(03): 665–685.

- Kasabov, N. (2001). Evolving fuzzy neural networks for online supervised/unsupervised online knowledge-based learning, *IEEE Transactions Systems, Man and Cybernetics* **31**(6): 902–918.
- Kasabov, N. e Filev, D. (2006). Evolving intelligent systems: Methods, learning & applications, *Proceedings of 2006 International Symposium on Evolving Fuzzy Systems – EFS'06*, IEEE Press, Ambleside, Lake District, UK, pp. 08–18.
- Kasabov, N. e Song, Q. (2002). DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time series prediction, *IEEE Transactions on Fuzzy Systems* **10**(02): 144–154.
- Kaymak, U. e Setnes, M. (2000). Extended fuzzy clustering algorithms, *Research paper*, Erasmus Research Institute of Management (ERIM).
- Klir, G. e Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Application*, 1ª edição, Prentice Hall, Upper Saddle River, NJ, EUA.
- Kohonen, T. (2001). *Self-Organizing Maps*, Springer Series in Information Sciences, 3ª edição, Springer-Verlag.
- Koza, J. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, Massachusetts, EUA.
- Kushner, H. e Clark, D. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, 1ª edição, Springer, New York.
- Lages, E. (2003). *Espaços Métricos*, Projeto Euclides, Associação Instituto Nacional de Matemática Pura e Aplicada, RJ, Brasil.
- Lee, M., Lee, S.-Y. e Park, C. (1995). Neurofuzzy controller design using neurofuzzy identifier, *International Journal of Approximate Reasoning* **13**(4): 269–285.
- Leng, G., McGinnity, T. e Prasad, G. (2005). An approach for on-line extraction of fuzzy rules using a self-organising fuzzy neural network, *Fuzzy Sets and Systems* **150**(2): 211–243.
- Lima, E., Ballini, R. e Gomide, F. (2006). Modelagem evolucionária participativa nebulosa, *XVI Congresso Brasileiro de Automática – CBA'06*, Salvador, BA, Brasil, pp. 2519–2524.

- Lima, E., Ballini, R. e Gomide, F. (2007). Modelagem de sistemas utilizando aprendizado participativo, *VIII Congresso Brasileiro de Redes Neurais – CBRN'07*, Florianópolis, SC, Brasil, pp. 2619–2624.
- Lima, E., Gomide, F. e Ballini, R. (2006). Participatory evolving fuzzy modeling, *Proceedings of the International Symposium on Evolving Fuzzy Systems – EFS'06*, IEEE Press, Ambleside, Lake District, UK, pp. 36–41.
- Lin, F., Lin, C. e Shen, P. (2001). Self-constructing fuzzy neural network speed controller for permanent-magnet synchronous motor drive, *IEEE Transactions on Fuzzy Systems* **9**(5): 751–759.
- Michaelis (1998). *Moderno Dicionário da Língua Portuguesa*, Editora Melhoramentos, São Paulo, SP, Brasil.
- Nelles, O. (2000). *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*, 1ª edição, Springer, Heidelberg, Alemanha.
- Nguyen, H. e Sugeno, M. (1998). *Fuzzy Systems: Modeling and Control*, The Handbooks of Fuzzy Sets Series, Kluwer Academic Publishers, Norwell, MA, EUA.
- Pal, N. e Chakraborty, D. (2000). Mountain and subtractive clustering method: Improvements and generalizations, *International Journal of Intelligent Systems* **15**(4): 329–341.
- Pedrycz, W. e Gomide, F. (1998). *An Introduction to Fuzzy Sets: Analysis and Design*, 1ª edição, MIT Press, Massachusetts, EUA.
- Pedrycz, W. e Gomide, F. (2007). *Fuzzy Systems Engineering: Toward Human-Centric Computing*, 1ª edição, Wiley-IEEE Press, EUA.
- Robbins, H. e Monro, S. (1951). A stochastic approximation method, *The Annals of Mathematical Statistics* **22**(3): 400–407.
- Roubens, M. (1982). Fuzzy clustering algorithms and their cluster validity, *European Journal of Operational Research* **10**(3): 294–301.
- Rumelhart, D., Hinton, G. e Williams, R. (1986). *Learning internal representation by error propagation*, 1ª edição, Nature, London, UK.
- Saridis, G. (1974). Stochastic approximation methods for identification and control – A survey, *IEEE Transactions on Automatic Control* **19**(6): 798–809.

- Silva, L. (2003). *Aprendizagem participativa em agrupamento nebuloso de dados*, Dissertação de Mestrado, FEEC–Unicamp, Campinas, Brasil.
- Silva, L., Gomide, F. e Yager, R. (2005). Participatory learning in fuzzy clustering, *Proceedings of the 14th IEEE International Conference on Fuzzy Systems, FUZZ-IEEE*, Reno, Nevada, EUA, pp. 857–861.
- Sussner, P. e Valle, M. (2007). *Morphological and Certain Fuzzy Morphological Associative Memories with Applications in Classification and Prediction*, Vol. 67 of *Computational Intelligence Based on Lattice Theory, Series: Studies in Computational Intelligence*, Kaburlasos, V. e Ritter, G. (Eds.), Springer Verlag.
- Takagi, T. e Sugeno, M. (1985). Fuzzy identification of systems and its application to modeling and control, *IEEE Transactions on Systems, Man and Cybernetics* **15**(1): 116–132.
- Wang, I., Chong, E. e Kulkarni, S. (1996). Equivalent necessary and sufficient conditions on noise sequences for stochastic approximation algorithms, *Advances in Applied Probability* **28**(3): 784–801.
- Wang, I.-J. e Chong, E. (1998). A deterministic analysis of stochastic approximation with randomized directions, *IEEE Transactions on Automatic Control* **43**(12): 1745–1749.
- Wasan, M. (1969). *Stochastic Approximation*, 1^a edição, Cambridge University Press, Cambridge, UK.
- Wellstead, P. e Zarrop, M. (1995). *Self-tuning systems: Control and Signal Processing*, 1^a edição, John Wiley & Sons, NY, EUA.
- Werbos, P. (1990). Backpropagation through time: What it does and how to do it, *Proceedings of the IEEE Conference on Neural Networks* **78**(10): 1550–1560.
- Yager, R. (1990). A model of participatory learning, *IEEE Transactions on Systems, Man and Cybernetics* **20**(05): 1229–1234.
- Yager, R. (2004). Participatory learning: A paradigm for more human like learning, *2004 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE*, Budapest, Hungary, pp. 79–84.
- Yager, R. e Filev, D. (1994). Generation of fuzzy rules by mountain clustering, *Journal of Intelligent and Fuzzy Systems* **02**(03): 209–219.

- Yen, J., Wang, L. e Gillespie, C. (1998). Improving the interpretability of TSK fuzzy models by combining global learning and local learning, *IEEE Transactions on Fuzzy Systems* **6**(4): 530–537.
- Young, P. C. (1984). *Recursive Estimation and Time-Series Analysis: An Introduction*, Springer-Verlag, NY, EUA.
- Zadeh, L. (1965). Fuzzy sets, *Information and Control* **8**: 338–353.
- Zadeh, L. (1973). Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Transactions on Systems, Man, and Cybernetics* **3**: 28–44.