

MODELO CONSTRUTIVO BASEADO EM REGRAS PARA PREVISÃO DE SÉRIES TEMPORAIS

IVETTE LUNA*, SECUNDINO SOARES*, ROSANGELA BALLINI†

**Universidade Estadual de Campinas – UNICAMP
Departamento de Engenharia de Sistemas – DENNIS
Faculdade de Engenharia Elétrica e de Computação – FEEC
13081-970 Campinas, SP, Brasil*

*†Universidade Estadual de Campinas – UNICAMP
Departamento de Teoria Econômica – DTE
Instituto de Economia – IE
13083-857 Campinas, SP, Brasil*

Emails: iluna@cose.fee.unicamp.br, dino@cose.fee.unicamp.br, ballini@eco.unicamp.br

Abstract— This paper suggests a constructive fuzzy rule based system based on Takagi-Sugeno system. The parameters and structure are determined by the traditional Expectation Maximization algorithm. The learning process is composed by two steps. In first an initial rule based is adjusted using the Expectation Maximization algorithm. In second step the initial rule based is modified and the structure is determined in a constructive fashion, via adding and pruning operators. The constructive learning process reduces model complexity and defines automatically the structure of the system, providing an efficient time series model. The model proposed is applied for building an economical time series model. Simulation results show the good performance of the approach, specially for a multi-step ahead prediction.

Keywords— Fuzzy rule based system, constructive learning, EM algorithm, economical time series prediction.

Resumo— Este artigo apresenta um sistema de aprendizado construtivo baseado em regras nebulosas, do tipo Takagi-Sugeno. Para o ajuste dos parâmetros e determinação da estrutura do sistema nebuloso é proposto um método construtivo baseado no algoritmo de Maximização da Verossimilhança. Esse processo de aprendizado é composto por duas etapas. Na primeira, uma base de regras inicial é ajustada usando o algoritmo de otimização da Maximização da Verossimilhança. Na segunda etapa a base de regras inicial é modificada, sendo a estrutura do modelo definida por meio de um aprendizado construtivo offline e de operadores de adição e eliminação de regras nebulosas. O aprendizado construtivo tem como vantagens, a redução da complexidade do modelo e a definição automática da estrutura do sistema, fornecendo assim, um modelo eficiente de série temporal. O desempenho do modelo proposto é verificado por meio da aplicação do sistema para a previsão de uma série temporal econômica. Os resultados de simulação mostram o desempenho satisfatório da abordagem proposta, especialmente na previsão vários passos à frente.

Palavras-chave— Sistema baseado em regras nebulosas, aprendizado construtivo, algoritmo EM, previsão de séries temporais econômicas.

1 Introdução

Modelos de séries temporais são amplamente empregados em problemas de predição de valores futuros, simulação ou simplesmente modelagem de dados (Brockwell and Davis, 2002). Muitos desses modelos são baseados na metodologia clássica de Box & Jenkins, a qual assume relação linear entre as variáveis do sistema (Box et al., 1994). A desvantagem deste tipo de modelagem é a perda de informação, dado que é considerado apenas relações lineares entre as variáveis que descrevem o problema, resultando em possível perda de informação.

Diante disso, o interesse por modelos de previsão não lineares, tais como redes neurais e sistemas nebulosos, vêm crescendo nas últimas décadas (Maier and Dandy, 2000). Embora estes modelos apresentem um processo de aprendizado mais complexo, eles mostram vantagens interessantes em termos de interpretabilidade, eficiência, representação do conhecimento e adaptação (Mitra and

Hayashi, 2000). No entanto, um problema ainda em estudo, é a seleção da estrutura adequada do modelo de previsão.

Como contribuição, este trabalho sugere um modelo baseado em regras nebulosas, com aprendizado offline e construtivo, que define de forma automática a estrutura do modelo de previsão. O modelo é baseado em sistemas do tipo Takagi-Sugeno (T-S), sendo construído em duas fases.

Na primeira fase, o modelo é inicializado com duas regras nebulosas do tipo Se <antecedente> Então <conseqüente>. A partir de um conjunto de dados entrada-saída, os parâmetros das duas regras nebulosas são ajustados usando o algoritmo tradicional de Maximização da Verossimilhança (EM) (Jacobs et al., 1991). Na segunda fase, a estrutura inicial é modificada e os parâmetros são ajustados utilizando um algoritmo construtivo baseado no algoritmo EM. Por meio de operadores de geração (ou adição) e eliminação (ou punição) de regras a estrutura da rede é modificada durante esta fase. Estes operadores são aplicados de

acordo com critérios previamente definidos, de tal forma que os dados do conjunto de treinamento estejam contidos em regiões ativas que são representadas pelas regras nebulosas. Isto é possível pois a base de regras define, de forma implícita, uma partição do espaço de entrada em sub-regiões, sendo um modelo local alocado a cada uma dessas sub-regiões. Com a modificação da base de regras durante o aprendizado, altera-se também a partição do espaço de entrada. Assim, o algoritmo proposto determina o número de regras nebulosas e realiza o ajuste dos parâmetros de forma simultânea.

O desempenho do modelo proposto é verificado usando dados reais de uma série econômica de taxa de desemprego trimestral dos Estados Unidos. Estes dados foram utilizados em (Montgomery et al., 1998) e em (Tsay, 2005) para a construção e avaliação de modelos lineares e não lineares de regressão. Os resultados obtidos com o modelo construtivo, para a previsão de um até cinco passos à frente, são comparados com os obtidos em (Tsay, 2005), mostrando a eficiência do modelo proposto.

Este artigo está organizado como segue. A Seção 2 descreve a estrutura geral do modelo baseado em regras nebulosas (**C-FSM**). A Seção 3 detalha o algoritmo de aprendizado construtivo utilizado. Os resultados de simulação são apresentados na Seção 4. Finalmente, conclusões e trabalhos futuros são descritos na Seção 5.

2 Estrutura geral do modelo

A estrutura do sistema proposto é composta por um conjunto de M regras nebulosas de primeira ordem, do tipo Takagi-Sugeno (TS) (Takagi and Sugeno, 1985), como ilustra a Figura 1.

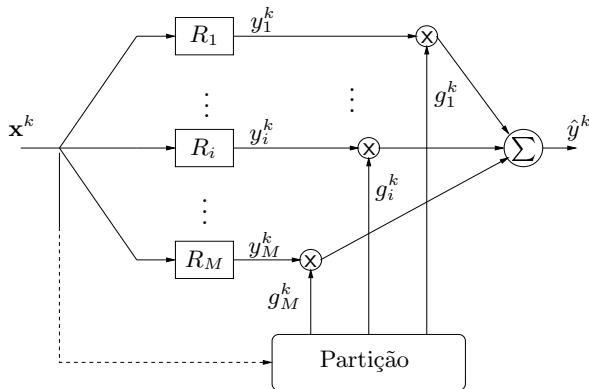


Figura 1: Estrutura do sistema baseado em regras nebulosas.

Seja $\mathbf{x}^k = [x_1^k, x_2^k, \dots, x_p^k] \in \mathbb{R}^p$ o vetor de entrada no instante k , $k \in \mathbb{Z}_0^+$; $\hat{y}^k \in \mathbb{R}$ a saída do modelo, para a entrada \mathbf{x}^k . A base de regras representada na estrutura do sistema, produz uma partição do espaço de entrada definida pelos ante-

cedentes das M regras nebulosas. Cada sub-região gerada, é definida por um centro $\mathbf{c}_i |_{(1 \times p)}$ e por uma matriz de covariância $\mathbf{V}_i |_{(p+1) \times (p+1)}$. Assim, cada vetor de entrada tem um grau de pertinência associado a cada uma das sub-regiões geradas pela base de regras.

O grau de pertinência de cada padrão \mathbf{x}^k é representado por $g_i(\mathbf{x}^k) \in [0, 1]$, calculado da seguinte forma:

$$g_i(\mathbf{x}^k) = g_i^k = \frac{\alpha_i \cdot P[i | \mathbf{x}^k]}{\sum_{q=1}^M \alpha_q \cdot P[q | \mathbf{x}^k]} \quad (1)$$

sendo α_i coeficientes positivos que satisfazem a condição:

$$\sum_{i=1}^M \alpha_i = 1 \quad (2)$$

Estes coeficientes também podem ser considerados como índices indiretos de relevância para cada regra gerada durante o treinamento, de tal forma que, quanto mais alto seja o valor de α_i , para $i = 1, \dots, M$, maior será a importância da i -ésima regra nebulosa. Além disso, a probabilidade $P[i | \mathbf{x}^k]$ na Eq.(1) é definida como:

$$P[i | \mathbf{x}^k] = \left(\frac{1}{(2\pi)^{p/2} \det(\mathbf{V}_i)^{1/2}} \right) \times \left(\exp \left\{ -\frac{1}{2} (\mathbf{x}^k - \mathbf{c}_i) \mathbf{V}_i^{-1} (\mathbf{x}^k - \mathbf{c}_i)^T \right\} \right) \quad (3)$$

na qual \mathbf{V}_i e \mathbf{c}_i são a matriz de covariância e centro associados a i -ésima regra, respectivamente. A Eq.(3) representa a probabilidade de escolher a regra i , dado que apenas a entrada \mathbf{x}^k é conhecida, assim como os parâmetros \mathbf{V}_i e \mathbf{c}_i . Observa-se que $P[i | \mathbf{x}^k]$ é uma função de densidade de probabilidade condicional, com centros em \mathbf{c}_i e matriz de covariância \mathbf{V}_i , que pode ser interpretada como a probabilidade do vetor de entrada \mathbf{x}^k pertencer à região ativa da i -ésima regra.

Além disso, da Eq. (1), observa-se que g_i^k satisfaz a condição:

$$\sum_{i=1}^M g_i^k = 1 \quad (4)$$

a qual é necessária para considerar g_i^k como uma função de pertinência (Pedrycz and Gomide, 1998).

A saída do sistema \hat{y}^k , é calculada como uma combinação não linear de modelos locais com saídas y_i^k ponderados pelos seus respectivos graus de pertinência g_i^k , ou seja:

$$\hat{y}^k = \sum_{i=1}^M g_i^k y_i^k \quad (5)$$

sendo, neste caso, a saída de cada regra i , y_i^k , definida como:

$$y_i^k = \phi^k \times \theta_i^T \quad (6)$$

na qual $\phi^k = [1 \ x_1^k \ x_2^k \ \dots \ x_p^k]$ é o vetor que contém os dados de entrada \mathbf{x}^k e o termo constante; e $\theta_i = [\theta_{i0} \ \theta_{i1} \ \dots \ \theta_{ip}]$ é o vetor de coeficientes dos modelos locais lineares, de dimensão $1 \times (p + 1)$ para $i = 1, \dots, M$. Embora as Equações de (1)-(5), tenham considerado apenas uma única saída, o modelo pode ser estendido para múltiplas saídas.

Como pode ser observado, a dinâmica do modelo pode ser interpretada como um mecanismo de inferência nebulosa, sendo a base de regras composta por regras $R_i, i = 1, \dots, M$, do tipo **Se** <antecedentes> - **Então** <conseqüentes>. Por outro lado, o modelo proposto pode também ser interpretado como uma arquitetura modular, sendo que o mecanismo de inferência representa o sistema de controle para cada módulo ou especialista (Guo and Xu, 1999), e o módulo em si, é representado por cada um dos modelos lineares definidos. No entanto, a maior dificuldade quando se trabalha com este tipo de modelagem, é a definição automática do número de regras nebulosas, M , necessárias para cada aplicação. Este trabalho propõe um algoritmo de aprendizado construtivo, chamado de aprendizado construtivo offline (**C-FSM**), o qual determina, de forma automática, o número de regras nebulosas na estrutura do modelo durante a fase de treinamento. Esta proposta de aprendizado construtivo é descrita a seguir.

3 Aprendizado Construtivo Offline

O processo de aprendizado construtivo **C-FSM** é composto de duas etapas: inicialização e adaptação da estrutura do modelo. Estas duas etapas são detalhadas a seguir.

3.1 Primeira Etapa: Inicialização

Inicialmente, uma estrutura composta por duas regras nebulosas é construída. Cada uma dessas regras é composta por parâmetros que definem os antecedentes: centro, matriz de covariância, índice de relevância; e os parâmetros relacionados ao vetor de coeficientes dos modelos locais, que definem o conseqüente. Cada um desses parâmetros é inicializado da seguinte forma: o centro é igual a um vetor de entrada \mathbf{x}^k ; a matriz de covariância é inicializada com $\mathbf{V}_i = 10^{-4}\mathbf{I}$, sendo \mathbf{I} a matriz identidade; o índice de relevância é igual a $\alpha_i = 1/M$ sendo M o número de regras; e, os parâmetros θ_i assumem valores aleatórios no intervalo $[0, 1]$.

Para o ajuste desses parâmetros, neste artigo, é proposto o algoritmo de Maximização da Verossimilhança, ou algoritmo EM (*Expectation Maximization*). Este algoritmo foi proposto para otimização de modelos de misturas de especialistas em (Jacobs et al., 1991). Posteriormente, a sua

aplicação foi estendida ao ajuste dos parâmetros de modelos de redes neurais.

Com base neste algoritmo de otimização, os parâmetros do modelo **C-FSM** são ajustados a partir de um conjunto de N dados de entrada-saída (\mathbf{x}^k, y^k) , chamado de conjunto de treinamento, durante uma seqüência iterativa de passos E (*Expectation*) e M (*Maximization*) (Jacobs et al., 1991). O objetivo do algoritmo EM, é a maximização da função de verossimilhança a cada passo M do algoritmo de aprendizado sendo necessário estimar as probabilidades a posteriori, $h_i^k, i = 1, \dots, M$, para o cálculo de valores ótimos dos parâmetros. Essas probabilidades a posteriori são aproximações das funções de probabilidade condicional definidas pela Eq.(3), sendo calculadas como:

$$h_i^k = \frac{\alpha_i P(i | \mathbf{x}^k) P(y^k | \mathbf{x}^k, \theta_i)}{\sum_{q=1}^M \alpha_q P(q | \mathbf{x}^k) P(y^k | \mathbf{x}^k, \theta_q)} \quad (7)$$

para $i = 1, \dots, M$, sendo que, estas aproximações são calculadas assumindo as saídas desejadas $y^k, k = 1, \dots, N$ como conhecidas. Além disso, a probabilidade condicional $P(y^k | \mathbf{x}^k, \theta_i)$ é definida por:

$$P(y^k | \mathbf{x}^k, \theta_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{[y^k - y_i^k]^2}{2\sigma_i^2}\right) \quad (8)$$

sendo θ_i e y_i^k o vetor dos coeficientes e a saída do modelo local relacionado à regra i , respectivamente. A variância σ_i^2 é calculada como:

$$\sigma_i^2 = \left(\sum_{k=1}^N h_i^k [y^k - y_i^k]^2 \right) / \sum_{k=1}^N h_i^k \quad (9)$$

Desta forma, enquanto no passo E as variáveis h_i^k são estimadas, a maximização da função objetivo e a estimação dos parâmetros são feitas durante o passo M. O processo continua até atingir a convergência, sendo os valores dos parâmetros do modelo ajustados pelas seguintes equações:

$$\alpha_i = \frac{1}{N} \sum_{k=1}^N h_i^k \quad (10)$$

$$\mathbf{c}_i = \left(\sum_{k=1}^N h_i^k \mathbf{x}^k \right) / \sum_{k=1}^N h_i^k \quad (11)$$

$$\mathbf{V}_i = \left(\sum_{k=1}^N h_i^k (\mathbf{x}^k - \mathbf{c}_i)' (\mathbf{x}^k - \mathbf{c}_i) \right) / \sum_{k=1}^N h_i^k \quad (12)$$

para $i = 1, \dots, M$, M o número de regras e $k = 1, \dots, N$, N o número de dados entrada-saída que compõem o conjunto de treinamento. Estas equações foram obtidas considerando as hipóteses de que a matriz de covariância \mathbf{V}_i é uma matriz diagonal e definida positiva, para evitar soluções infactíveis. A estimação de θ_i é obtida a partir da solução da equação:

$$\sum_{k=1}^N \frac{h_i^k}{\sigma_i^2} (y^k - \phi^k \times \theta_i) \cdot \phi^k = 0 \quad (13)$$

na qual σ_i é o desvio padrão associado a cada uma das saídas locais y_i^k , $i = 1, \dots, M$, sendo σ_i^2 a variância definida pela Eq.(9).

Após a inicialização do modelo, procede-se ao ajuste adaptativo o qual é baseado no algoritmo EM e em operadores de geração ou adição de novas regras nebulosas, assim como em operadores de punição ou eliminação, quando necessário.

3.2 Segundo Estágio: Adaptação

Como descrito na Seção 2, o problema de identificação do modelo de séries temporais é um problema sob constante estudo. Neste trabalho, é proposto o algoritmo de aprendizado construtivo baseado no algoritmo EM, de forma que, o número de regras nebulosas necessárias para construir os modelos de séries temporais seja determinado de forma automática e em paralelo à atualização dos parâmetros do modelo.

O algoritmo construtivo consiste na modificação da estrutura inicial descrita na subseção 3.1. Essas alterações são executadas durante o processo de aprendizado, mediante a aplicação de operadores de adição e de eliminação de regras nebulosas. Quando o modelo atual não satisfaz as condições desejadas uma nova regra é gerada. No caso em que uma das regras que compõem a estrutura atual do modelo não é mais relevante, esta é eliminada da estrutura. Tanto na adição como na eliminação de regras, a estrutura do modelo é atualizada e os parâmetros são recalibrados, tendo sempre em consideração as restrições e condições definidas pelas Equações (10)-(9).

Desta forma, o método de aprendizado construtivo capacita o sistema nebuloso a compensar degradações de desempenho provocadas por variação dos parâmetros do modelo ou outras perturbações devido à mudanças provocadas no sistema real.

Os critérios para a aplicação dos operadores de adição e de eliminação de regras nebulosas são descritos a seguir.

3.2.1 Geração de novas regras

A condição para a geração de uma nova regra no sistema é chamada de critério do antecedente. Este critério certifica-se de que a base de regras cobre todo o espaço de entrada, ou seja, os dados de entrada apresentados ao sistema devem pertencer a uma região ativa deste espaço, representada por alguma das regras nebulosas existentes. Assim, a máxima probabilidade do dado de entrada \mathbf{x}^k pertencer a pelo menos uma das M regras nebulosas deve ser superior a um limiar pré-definido.

Assumindo que os dados tenham uma distribuição normal, pode-se construir o intervalo de

confiança $[\mathbf{c}_i - z_\gamma \sqrt{\text{diag}(\mathbf{V}_i)}, \mathbf{c}_i + z_\gamma \sqrt{\text{diag}(\mathbf{V}_i)}]$, no qual \mathbf{c}_i é o centro e $\text{diag}(\mathbf{V}_i)$ é a diagonal principal da matriz \mathbf{V}_i , para um dado coeficiente de confiança $\gamma\%$. Neste artigo, foi considerado um coeficiente de confiança $\gamma = 72,86\%$, tal que $P(-z_\gamma < Z < z_\gamma) = 72,86\%$, com $Z \sim N(0,1)$. Assim, para este coeficiente de confiança, a máxima probabilidade da i -ésima regra estar ativa para um dado vetor de entrada \mathbf{x}^k é de 0,1357, ou seja:

$$\max (P[i | \mathbf{x}^k])_{i=1, \dots, M} > 0,1357 \quad (14)$$

Se a condição definida pela Eq. (14) não é satisfeita por um ou mais dados do conjunto de treinamento, significa que estes dados possuem uma probabilidade de pertencer a alguma região ativa coberta pela base de regras, inferior ao desejado. Denotando por Ω o conjunto contendo todos os dados entrada-saída que não satisfazem a Eq. (14), uma nova regra será inserida à estrutura do sistema, ou seja,

$$M = M + 1 \quad (15)$$

e o centro desta nova regra \mathbf{c}_M é inicialmente estimado como:

$$\mathbf{c}_M = \frac{1}{N_\Omega} \sum_{t \in \Omega} \mathbf{x}^t \quad (16)$$

sendo N_Ω o número de dados $[\mathbf{x}^t y^t]$ em Ω . Além disso, os demais parâmetros associados a essa nova regra são inicializados da seguinte forma:

- $\sigma_M = 1, 0$;
- $\theta_M = [\bar{y} \ 0 \ \dots \ 0]_{1 \times p+1}$, sendo \bar{y} o valor médio das saídas desejadas pertencentes ao conjunto Ω .

Os parâmetros α_i e \mathbf{V}_i , para $i = 1, \dots, M$, devem ser re-inicializados de acordo com:

- $\mathbf{V}_i = 10^{-4} \mathbf{I}$, sendo $\mathbf{I}_{p \times p}$ a matriz identidade;
- $\alpha_i = 1/(M)$.

Após a inserção da nova regra, a estrutura do sistema é re-ajustada durante algumas iterações EM. Como α_i indica a relevância de cada regra nebulosa, a sua re-inicialização é necessária para que todas as regras tenham o mesmo peso, fazendo com que o ajuste dos parâmetros seja não-tendencioso, obtendo assim uma nova partição do espaço de entrada. A re-inicialização de \mathbf{V}_i é necessária já que o processo de re-ajuste definirá novos valores de dispersão para cada regra nebulosa, de acordo com a nova partição gerada, ou seja, considerando neste processo a nova regra inserida na estrutura do modelo.

3.2.2 Eliminação ou Poda

O operador de eliminação ou poda consiste em retirar regras nebulosas da estrutura do modelo, que não estejam contribuindo de forma relevante para o cálculo da saída y^k . Nota-se na Eq. (10), que o valor para α_i a cada iteração é uma soma ponderada das estimações a posteriori dos graus de pertinência sob o conjunto de treinamento. Assim, quanto mais a i -ésima regra é ativa, maior será o valor do seu respectivo α_i .

Esta característica é importante para a eliminação ou poda de regras nebulosas, já que, se for necessário eliminar alguma regra, aquela com menor α_i será uma boa candidata. Para tal, um limiar para este parâmetro deve ser definido previamente, de tal forma que, toda regra com $\alpha_i < \alpha_{min}$ a cada iteração, será eliminada da estrutura do sistema. Neste trabalho, foi definido um limiar $\alpha_{min} = 0,01$. Assim, a regra será eliminada se o valor de α for menor que 0,01.

Após a eliminação de alguma regra, o modelo será reajustado durante algumas iterações EM por meio das Equações definidas na Seção 2 e na Subseção 3.1.

O processo global de aprendizado é finalizado após atingir a convergência, e quando não exista mais mudanças na estrutura do modelo.

A utilização da função de verossimilhança como função objetivo, em conjunto com o algoritmo EM é de interesse neste trabalho, já que este não só diminui o erro quadrático médio, mas também a variância do erro, aumentando assim, as possibilidades de obter melhores resultados.

O primeiro estágio de inicialização descrito para esta abordagem de aprendizado offline é necessário, já que, o algoritmo EM é sensível à inicialização. Como o algoritmo de aprendizado construtivo offline é baseado nas Equações (10)-(9), as quais são derivadas do algoritmo de otimização EM, uma boa inicialização aumentará as possibilidades de obter modelos eficientes no final do aprendizado.

4 Resultados de Simulação

Para analisar o desempenho do sistema nebuloso com aprendizado construtivo **C-FSM** para previsão de séries temporais, considerou-se uma série econômica, composta pela taxa trimestral de desemprego dos Estados Unidos, com histórico a partir do primeiro trimestre de 1948 (1948-I), até o quarto trimestre de 1993 (1993-IV). Esta série, sazonalmente ajustada, é calculada como a média ponderada das taxas de desemprego mensais para cada trimestre, sendo analisada de forma detalhada em (Montgomery et al., 1998).

Algumas características dessa série são: a série possui um comportamento cíclico altamente relacionado aos ciclos econômicos dos Estados Unidos; e a série apresenta crescimento acelerado da

taxa de desemprego, período que finaliza em picos elevados, os quais são alternados com quedas menos acentuadas. Estas características sugerem que a dinâmica da série é de natureza não linear.

Os resultados obtidos pelo **C-FSM** são comparados aos resultados obtidos em (Tsay, 2005) para o modelo linear ARIMA (*linear univariate autoregressive integrated moving average model*) e para os modelos não lineares MSA (*Markov switching autoregressive model*) e TAR (*threshold autoregressive model*).

O primeiro passo para a construção de um modelo de previsão é a seleção de entradas. Neste sentido, e para comparar o desempenho dos modelos, as entradas utilizadas neste trabalho são as mesmas utilizadas pelos modelos não lineares TAR e MSA em (Tsay, 2005). Assim, foi necessário trabalhar com a primeira diferença da série $\Delta x_t = x_t - x_{t-1}$, sendo as entradas iguais a Δx_{t-1} e Δx_{t-2} .

Após definir as entradas do modelo, a estrutura final é definida de acordo com os critérios descritos na Seção 2 e Seção 3. O modelo foi ajustado utilizando apenas dados do histórico desde 1948-I até 1968-II, dando um total de 79 dados entrada-saída no conjunto de treinamento ou calibração. O período de previsão é composto por dados a partir de 1968-III até 1993-IV, dando um total de 102 dados. Como descrito na Subseção 3.2.2, o valor de α_{min} é igual a 0,01, sendo o aprendizado finalizado após 40 épocas. Observou-se uma variação no número de regras entre 2 e 4, sendo a estrutura final composta por 4 regras nebulosas.

A seguir, o modelo estimado foi avaliado, com o intuito de verificar a qualidade do ajuste. Neste trabalho, a verificação foi feita via uma análise dos resíduos estimados para o conjunto de treinamento. Assim, se o modelo for adequado, os resíduos estimados deverão ser aproximadamente não-correlacionados. Dois testes de adequação foram usados: o teste de autocorrelação residual e o teste de correlação cruzada entre os resíduos estimados e a série temporal. A Figura 2 ilustra ambos testes, sendo que as linhas pontilhadas indicam os limites $\pm 2/\sqrt{n}$, os quais fornecem uma indicação geral de possível quebra do comportamento de ruído branco, com n igual ao número de graus de liberdade do modelo obtido. Ambos testes mostram que o modelo ajustado é adequado, pois os resíduos estimados são não correlacionados.

Os modelos foram comparados utilizando o erro quadrático médio como métrica de desempenho, na previsão de um até cinco passos à frente, a partir de 1968-III. A Tabela 1 apresenta os erros obtidos para cada modelo de previsão.

Observa-se que, os modelos de regressão linear e não linear apresentaram um melhor desempenho na previsão um passo à frente, e um desempenho similar na previsão dois passos à frente. O modelo

C-FSM obteve melhor desempenho na previsão até 3, 4 e 5 passos à frente, mostrando assim, a sua eficiência e robustez na previsão até vários passos à frente. A Figura 3 ilustra a previsão do **C-FSM** até 5 passos à frente.

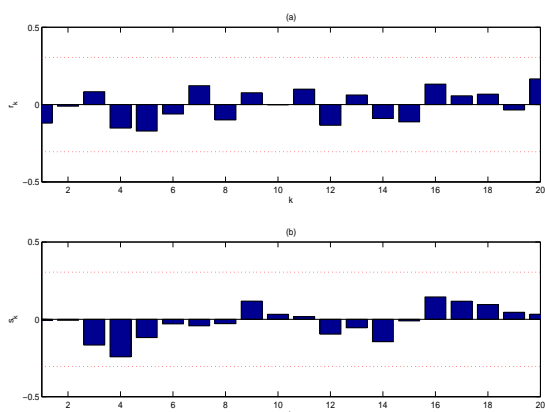


Figura 2: Funções de autocorrelações estimadas dos resíduos e função de correlação cruzada estimada para a série em estudo.

Tabela 1: Desempenho dos modelos de previsão.

Modelo	1	2	3	4	5
C-FSM	0,55	0,58	0,60	0,61	0,59
Linear	0,08	0,31	0,67	1,13	1,54
MSA	0,10	0,43	0,94	1,64	2,48
TAR	0,08	0,32	0,66	1,11	1,59

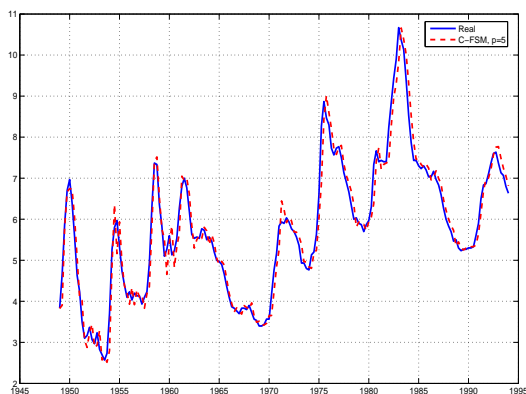


Figura 3: Previsão até 5 passos à frente da taxa de desemprego trimestral dos Estados Unidos.

5 Conclusões

Este trabalho sugere um modelo baseado em regras nebulosas com aprendizado construtivo para previsão de séries temporais. A estratégia de aprendizado construtivo é baseada no algoritmo de otimização EM, e em operadores de geração e eliminação de regras da estrutura do modelo. Estes operadores são aplicados durante o aprendizado com o objetivo de definir, de forma automática, a estrutura do modelo nebuloso. A análise

do desempenho do modelo proposto foi verificada usando uma série econômica de taxa de desemprego trimestral dos Estados Unidos. O modelo construtivo foi aplicado na previsão de um até cinco passos à frente e os resultados foram comparados com modelos de regressão linear e não linear. Os resultados de simulação mostraram que o sistema nebuloso proposto fornece modelos de previsão adequados e eficientes para previsão vários passos à frente.

Agradecimentos

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo auxílio.

Referências

- Box, G., Jenkins, G. and Reinsel, G. C. (1994). *Time Series Analysis, Forecasting and Control*, 3rd edn, Holden Day, Oakland, California, EUA.
- Brockwell, P. J. and Davis, R. A. (2002). *Introduction to time series and forecasting*, 2nd edn, Springer.
- Guo, P. and Xu, L. (1999). Relationship between Mixture of Experts and Ensemble Neural Networks, *Proceedings of the 6th International Conference on Neural Information Processing 1*: 246–250.
- Jacobs, R., Jordan, M., Nowlan, S. and Hinton, G. (1991). Adaptive Mixture of Local Experts, *Neural Computation* **3**(1): 79–87.
- Maier, H. and Dandy, G. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, *Environmental Modelling & Software* **15**: 101–124.
- Mitra, S. and Hayashi, Y. (2000). Neuro-fuzzy rule generation: survey in soft computing framework, *IEEE Transactions on Neural Networks* **11**(3): 748–768.
- Montgomery, A. L., Zarnowitz, V., Tsay, R. S. and Tiao, G. C. (1998). Forecasting the U.S. Unemployment Rate, *Journal of the American Statistical Association* **93**(442): 478–493.
- Pedrycz, W. and Gomide, F. (1998). *An Introduction to Fuzzy Sets: Analysis and Design*, MIT Press, Cambridge, MA.
- Takagi, T. and Sugeno, M. (1985). Fuzzy Identification of Systems and Its Applications to Modeling and Control, *IEEE Transactions on Systems, Man and Cybernetics* (1): 116–132.
- Tsay, R. S. (2005). *Analysis of Financial Time Series*, 2nd edn, Wiley & Sons.