

MODELAGEM DE SISTEMAS UTILIZANDO APRENDIZADO PARTICIPATIVO

ELTON LIMA*, ROSANGELA BALLINI†, FERNANDO GOMIDE*

**Faculdade de Engenharia Elétrica e Computação
Universidade Estadual de Campinas
Campinas, São Paulo, Brasil 13083-852*

*†Instituto de Economia
Universidade Estadual de Campinas
Campinas, São Paulo, Brasil 13083-857*

Emails: elton@cose.fee.unicamp.br, ballini@eco.unicamp.br, gomide@dca.fee.unicamp.br

Abstract— This paper introduces an approach to develop evolving fuzzy rule-based models based on the idea of participatory learning. Participatory learning is a mean to learn and revise beliefs based on what is already known or believed. Participatory learning naturally induces unsupervised dynamic fuzzy clustering algorithms and provides an effective alternative to construct evolving functional fuzzy models and adaptive fuzzy systems. Evolving participatory learning is used for short-term load forecasting. The results are compared with eTS, an evolving modeling technique that uses the notion of potential to dynamically cluster data. The results show the usefulness of the model proposed to time series forecasting.

Keywords— Fuzzy Clustering Algorithms, Participatory Learning, Time Series Forecasting.

Resumo— Este artigo propõe uma abordagem para o desenvolvimento de modelos adaptativos baseados em agrupamento de dados utilizando aprendizado participativo. O aprendizado participativo é analisado como um algoritmo de agrupamento nebuloso não supervisionado e inerentemente dinâmico, fornecendo uma alternativa eficiente na construção de modelos nebulosos evolutivos e adaptativos. O desempenho do aprendizado participativo evolutivo é analisado em um problema de previsão de série de carga a curto prazo. O modelo proposto é comparado com a técnica de modelagem evolucionária baseada na noção de potencial para agrupar dados dinamicamente. Os resultados mostram que o modelo proposto é eficiente para previsão de séries temporais.

Palavras-chave— Algoritmos de Agrupamento Nebuloso, Aprendizado Participativo, Previsão de Séries Temporais.

1 Introdução

A descrição de fenômenos relacionados a acontecimentos naturais representa um problema complexo, devido principalmente às características não lineares, incertezas e imprecisões. Nesse sentido, a modelagem baseada em regras nebulosas representa uma alternativa que visa levar em conta características de difícil interpretação (Takagi and Sugeno, 1985). No entanto, ao descrever um determinado fenômeno utilizando regras, deve-se buscar maneiras de adequar os parâmetros ao problema, bem como decidir a forma na qual os mesmos serão modificados quando uma nova informação é coletada. Assim, busca-se uma forma de construir funções de pertinência, termos linguísticos e parâmetros, atualizando-os a cada instante.

Até o início da década de 90, as principais abordagens para ajuste de parâmetros eram baseadas em métodos do tipo gradiente e, mais recentemente, algoritmos genéticos (Anderson et al., 1988). Durante os anos 90, novas idéias para extração de conhecimento em conjuntos de dados foram desenvolvidas com o objetivo de identificar um modelo capaz de descrever de forma eficiente o seu comportamento (Jang, 1993), (Chiu, 1994), (Angelov and Guthke, 1997). Nesse período, dentre outras técnicas, destacam-se os algoritmos de agrupamento e modelagens baseadas em problemas de otimização (Jang, 1993), (Chiu, 1994).

Em (Angelov, 2002) e (Angelov and Filev, 2004), um conceito de base de regras nebulosas adaptativa denominado modelo evolutivo Takagi-Sugeno (eTS) foi descrito. Nessa abordagem, uma função potencial é calculada e utilizada como critério para a atualização da base de regras. Essa atualização é feita de forma dinâmica, alterando uma das regras presentes ou adicionando alguma outra que melhor descreva o sistema. Para cada regra, um centro é definido e utilizado no cálculo das funções de pertinência do antecedente, sendo que no conseqüente são ajustadas funções lineares. Com isso, a alteração da regra é feita a partir da modificação do seu centro, também chamado de ponto focal, e dos respectivos parâmetros da função linear do conseqüente.

Diferentemente do eTS, em (Lima et al., 2006a) e (Lima et al., 2006b) foi proposta uma abordagem denominada aprendizado participativo evolutivo (APE) para a interpretação das novas informações e desenvolvimento de uma base de regras evolutiva. Essa abordagem combina a noção de aprendizado participativo, introduzida por Yager (Yager, 1990), com o algoritmo eTS, proposto por Angelov (Angelov, 2002), (Angelov and Filev, 2004). Assim, ao invés do cálculo de informações sobre a dispersão ou potencial dos dados, no APE os centros de grupos são atualizados a partir da combinação linear entre a nova informa-

ção e o ponto focal da regra mais significativa. Em (Lima et al., 2006b) os conseqüentes das regras nebulosas são funções lineares, sendo os parâmetros atualizados pelo método dos mínimos quadrados.

Neste artigo é implementada uma versão recursiva do mínimos quadrados para o ajuste dos parâmetros dos conseqüentes das regras nebulosas. Com isso, busca-se um algoritmo adaptativo no qual seja necessário um número menor de informações para a inicialização da base de regras, e para que a atualização desta seja feita e aprimorada de acordo com a obtenção de novas informações a cada instante. Além disso, é proposta também uma pequena formalização do comportamento do algoritmo APE.

O desempenho do algoritmo APE foi testado em um problema de previsão de séries temporais, representadas por dados de carga à curto prazo de uma empresa do setor elétrico brasileiro. Uma comparação dos resultados dos algoritmos APE e eTS é feita. Os resultados mostram que o algoritmo proposto é eficiente e a estrutura obtida é parcimoniosa.

2 Modelagem Baseada em Regras Nebulosas Evolutivas

Modelos funcionais nebulosos associados à uma base de regras são conhecidos como Takagi-Sugeno (TS), no qual os antecedentes são conjuntos nebulosos e os conseqüentes funções. Tomando essas funções como ajustes lineares de um conjunto de dados, tem-se a seguinte estrutura:

$$\mathcal{R}_i : \text{SE } x \Gamma_i \text{ ENTÃO } y_i = \gamma_{i0} + \sum_{j=1}^p \gamma_{ij} x_j$$

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

sendo \mathcal{R}_i , i -ésima regra nebulosa; c^k o número de regras no passo k ; $x \in [0, 1]^p$ o vetor de observações; y_i a saída da i -ésima regra; Γ_i o vetor de pertinência nebulosa dos antecedentes; e γ_{ij} os parâmetros dos conseqüentes.

A cada instante $k = 1, 2, \dots$, o espaço é dividido em c^k grupos que definem regiões nas quais modelos lineares são ajustados. Dessa forma, para cada grupo uma regra nebulosa é formada, sendo o conseqüente obtido a partir de um ajuste linear local, e a saída caracterizada por uma combinação desses c^k ajustes.

A contribuição da nova informação na saída de cada regra é dada pelo respectivo grau de ativação. Assim, definindo os conseqüentes das regras por funções de pertinência gaussianas, tem-se:

$$\bar{\mu}_{ij} = e^{-r|x_j^k - v_{ij}|^2} \quad (1)$$

sendo $j = 1, \dots, p$ a dimensão do espaço de entradas, r um parâmetro positivo que representa a zona de influência (raio) de cada modelo local, v_i o centro da regra $i = 1, \dots, c^k$ e $x^k \in [0, 1]^p$ a informação no instante $k = 1, 2, \dots$.

Com isso, seguindo (Angelov, 2002), o grau de ativação de cada regra pode ser dado por:

$$\mu_i = \prod_{j=1}^p \bar{\mu}_{ij} = e^{-r\|x^k - v_i\|^2} \quad (2)$$

A saída da base de regras deve ser tal que combine todas as saídas locais com os respectivos graus de ativação. Há diversas formas de obter essa combinação sendo que, no caso do eTS, a saída é calculada como:

$$y^k = \frac{\sum_{j=1}^{c^k} \mu_j y_j}{\sum_{j=1}^{c^k} \mu_j} \quad (3)$$

sendo c^k o número de regras após a k -ésima observação e y_j a saída da regra j . Para maiores detalhes ver (Angelov, 2002), (Angelov and Filev, 2004).

Para a inicialização da base de regras existem diversas técnicas, que dependem de como os parâmetros do conseqüente estão sendo calculados. Em modelos que utilizam ajuste global no conseqüente, como em (Lima et al., 2006a), a inicialização pode ser feita a partir de algum agrupamento prévio de um histórico das observações, gerando assim o número inicial de regras. Contudo, como mostrado nesse artigo, os parâmetros do conseqüente também podem ser obtidos de forma recursiva, e assim apenas uma informação inicial acerca do primeiro centro de grupo é necessária.

3 Modelagem Utilizando Aprendizado Participativo

O aprendizado participativo evolutivo (APE) representa uma modelagem baseada em regras nebulosas do tipo Takagi-Sugeno construídas de forma adaptativa e variável a cada passo. Assim como no eTS, o APE ajusta pertinências gaussianas nos antecedentes, sendo os parâmetros do conseqüente obtidos por mínimos quadrados.

Para a interpretação das observações e atualização dos antecedentes, o APE baseia-se na noção de aprendizado participativo (Yager, 1990). Diferentemente do eTS, que utiliza uma função potencial para decidir qual regra será modificada, o APE associa a cada informação uma medida de similaridade nebulosa. Dessa forma, a base de regras é ajustada a cada passo, atualizando a regra que possui maior grau de similaridade ou criando um nova regra.

Além disso, o aprendizado participativo possui um mecanismo que indica a presença de regras redundantes. Nesse caso, regras com centros muito similares, isto é, que geram graus de ativação muito próximos, podem ser substituídas por uma única regra equivalente.

Tomando $v_i^k \in [0, 1]^p$ como o centro da i -ésima regra no k -ésimo passo, o objetivo do

mecanismo de aprendizado participativo é atualizar essa variável a partir das observações reais x^k obtidas em cada instante $k = 1, \dots$.

Assim, dada uma estrutura inicial, a atualização é feita utilizando uma medida de similaridade nebulosa, $\rho_i^k \in [0, 1]$, e um índice de alerta, $a_i^k \in [0, 1]$, calculados para cada regra $i = 1, 2, \dots, c^k$ em cada instante $k = 1, 2, \dots$.

Enquanto a medida de similaridade mede o quanto o novo ponto é compatível com a estrutura atual da base de regras, o índice de alerta representa uma medida crítica capaz de detectar o grau com que novas informações são coerentes com o atual estágio do aprendizado.

Dado um parâmetro $\tau \in [0, 1]$, que define um limiar para o índice de alerta a_i^k , uma regra é criada caso $a_i^k > \tau$. Caso contrário, a regra que possui o maior grau de similaridade é atualizada de acordo com a igualdade:

$$v_i^{k+1} = v_i^k + G_i^k(x^k - v_i^k) \quad (4)$$

sendo:

$$G_i^k = \alpha \rho_i^k; \quad \rho_i^k = 1 - \frac{d_i^k}{p} \quad (5)$$

$\alpha \in [0, 1]$ é a taxa do aprendizado, $d_i^k = d(x^k, v_i^k)$ uma métrica e:

$$i = \arg \max_j \{\rho_j^k\} \quad (6)$$

com p a dimensão do espaço de observações.

Analogamente a (4), o índice de alerta a_i^k é dado por:

$$a_i^{k+1} = a_i^k + \beta(1 - \rho_i^{k+1} - a_i^k) \quad (7)$$

O valor de $\beta \in [0, 1]$ 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.

O centro v^k , chamado também de estado atual de aprendizado, é modificado na direção da nova observação x^k , sendo desejável que o tamanho do passo dependa tanto da medida de similaridade quanto do índice de alerta. Assim, o passo G_i^k em (4) deve ser maior quão maior for a similaridade e menor o alerta. Dentre as diversas formas de definir esse passo, (Yager, 2004) propôs uma igualdade do tipo:

$$G_i^k = \alpha(\rho_i^k)^{1-a_i^k} \quad (8)$$

Com isso, quando $a_i^k = 0$ tem-se $G_i^k = \alpha \rho_i^k$ representando a mesma igualdade (5), equivalente a um mecanismo sem alerta. Quanto maior for o alerta, menor será a influência do índice de compatibilidade na atualização, lembrando que, por construção, uma regra é atualizada somente se $a_i^k < \tau$.

A construção da base de regras é feita utilizando um procedimento similar ao eTS, descrito na seção 2. Para cada grupo uma regra é criada e

um modelo linear é ajustado usando o método de mínimos quadrados global ou recursivo.

O algoritmo APE possui um mecanismo que calcula a similaridade entre os centros, pois a atualização dada por (4) pode gerar centros muito próximos, e conseqüentemente saídas com grau de ativação semelhantes. Assim, após a atualização do centro, calcula-se a compatibilidade do mesmo em relação às outras regras com o objetivo de verificar se há informações redundantes.

Tomando um parâmetro $\lambda \in [0, 1]$ como limiar, caso a compatibilidade entre dois centros seja maior que esse parâmetro, as regras redundantes serão substituídas por uma única. Nesse caso, dentre as diversas estratégias possíveis, pode-se eliminar um dos centros, gerando uma única regra com o centro remanescente, ou mesmo substituir os dois centros por um outro formado pela média ponderada entre eles.

O cálculo da compatibilidade entre os centros é feito de forma similar à (5), ou seja:

$$\rho_{v_i}^k = 1 - \sum_{j=1}^p |v_i^k - v_j^k| \quad (9)$$

Assim, caso a condição:

$$\lambda_{v_i}^k = \beta(1 - \rho_{v_i}^k) \leq \lambda \quad (10)$$

seja satisfeita, a regra i é declarada como redundante e o seu centro é alterado. Para maiores detalhes do algoritmo APE ver (Lima et al., 2006b).

4 Dinâmica do Aprendizado Participativo Evolutivo

A atualização dos centros de grupo, bem como o ajuste do conseqüente da base de regras, caracterizam o comportamento geral do algoritmo APE, isto é, definem como o algoritmo irá processar as informações coletadas a cada instante. Analisar como essa dinâmica ocorre pode representar uma forma de entender como o algoritmo se comporta em relação ao espaço de observações.

Nesse sentido, o objetivo desta seção é descrever como os centros se comportam em relação a cada observação coletada e como se dá a influência no ajuste do conseqüente a cada instante.

4.1 Atualização dos centros

O algoritmo APE pode ser visto como uma busca em direção ao ótimo local representado pelo valor desconhecido a ser estimado. Nesse sentido, assume-se que cada x^k define a melhor estimativa possível para um determinado valor ótimo x^* , desconhecido devido às restrições reais, como por exemplo limite de precisão.

Partindo desse princípio, pode-se definir v^k como o ponto que caminhará em direção à estimativa atual, x^k , do valor ótimo x^* . Em outras

palavras, v^k representa o estado atual do aprendizado de x^* , sendo x^k a sua melhor estimativa.

Dessa forma, localmente a dinâmica do APE é tal que a cada instante otimiza uma função cujo ótimo está em x^k , sendo que o tamanho do passo é dado pela qualidade dessa estimativa, quantificada pelos índices de similaridade e alerta. Quanto melhor for essa estimativa x^k , mais próximo o centro v^{k+1} estará de x^k , pois supõe-se que menor é o erro nesse caso.

Sendo assim, o aprendizado definido em (4) pode ser visto como uma atualização do tipo:

$$v^{k+1} = v^k + \alpha^k d^k \quad (11)$$

onde $\alpha^k \in [0, 1]$ e $d^k \in [0, 1]^p$ são o passo e a direção no instante k , respectivamente.

Tomando a função quadrática em \mathbb{R}^p :

$$F(x) = \frac{1}{2}x^t A x - b^t x + c \quad (12)$$

tem-se o novo centro pertencente a uma reta que contém o ótimo de (12), para uma matriz A , um vetor b e um escalar c constantes.

Portanto, no algoritmo APE, a cada instante os centros são atualizados de acordo com um método do gradiente aplicado localmente, como mostrado no resultado a seguir.

Teorema 1 *Partindo da atualização (4), a cada instante $k = 1, 2, \dots$, o novo centro v^{k+1} se move na direção do gradiente de uma função quadrática na forma (12) para uma matriz $A \in \mathbb{R}^{p \times p}$, um vetor $b \in \mathbb{R}^p$ e um escalar c , fixos para todo k .*

Prova: Tomando $b = 0 \in \mathbb{R}^p$, $c = 0$ e A a matriz identidade, tem-se:

$$F(v) = \frac{1}{2}v^t v \quad \Rightarrow \quad \nabla F(v) = v$$

Considerando essa função quadrática centrada em x^k , isto é, $F(v) = \frac{1}{2}(x^k - v)^t(x^k - v)$ e $F(v^k) = F^k$, a atualização (4) representa:

$$v^{k+1} = v^k + \alpha^k \nabla_v F^k \quad (13)$$

isto é, para $d^k = \nabla_v F^k$ em (11), o novo centro v^{k+1} estará na direção do gradiente de F^k , sobre a reta dada pelos vetores x^k e v^k .

Definindo α^k a partir de (8), tem-se:

$$v^{k+1} = v^k + G^k \nabla_v F^k \quad (14)$$

obtendo a mesma igualdade dada em (4). \square

Como conseqüência, o estado do aprendizado no instante seguinte será sempre direcionado ao ótimo de F^k dado pela aproximação de x^* , sendo o tamanho do passo dependente do α^k .

Por construção, o APE possui um mecanismo de decisão caracterizado pelos índices de compatibilidade e alerta, definidos em (5)-(7). Deseja-se, portanto, que o índice de compatibilidade seja maior quanto menor for o erro entre x^k e x^* .

4.2 Atualização dos conseqüentes

A cada instante o conjunto de dados, $X^k = \{x^1 \dots x^k\}$, pode ser representado pela união de c^k conjuntos distintos, sendo c^k o número de regras no instante k . Dessa forma, tem-se:

$$X^k = \bigcup_{i=1}^{c^k} X_i \quad (15)$$

na qual X_i define o conjunto de pontos responsáveis pelo ajuste linear do conseqüente da regra $i = 1, \dots, c^k$, ou seja:

$$X_i = \left\{ x^k \left| \rho_i^k = \max_{j=1, \dots, c^k} \{\rho_j^k\} \right. \right\}, k = 1, 2, \dots \quad (16)$$

Em outras palavras, em cada regra os conseqüentes são ajustados utilizando os pontos que satisfazem a condição (6). Portanto, o conjunto X_i é formado pelas observações que ajustaram os parâmetros da regra $i = 1, 2, \dots, c^k$. Como a cada instante apenas a regra com maior índice de similaridade é atualizada, tem-se que:

$$X^i \cap X^j = \emptyset, \quad \forall i \neq j \quad (17)$$

Dessa forma, cada conseqüente define um ajuste local responsável pela descrição linear das informações de cada grupo, sendo a saída (3) da base de regras uma composição de todas as saídas y_i . Segundo a linearização local dada pela regra i no instante k definida por:

$$y_i^k = \gamma_{i0} + \sum_{j=1}^n \gamma_{ij} x_j \quad (18)$$

tem-se que:

$$\frac{\partial y_i^k}{\partial x_j} = \gamma_{ij} \quad (19)$$

$i = 1, \dots, c^k$ e $j = 1, \dots, p$ sendo p a dimensão do vetor de entradas.

Em cada direção, a saída (3) define uma combinação das inclinações de cada ajuste local, isto é:

$$\frac{\partial y^k}{\partial x_j} = \sum_{i=1}^{c^k} \mu_i \gamma_{ij} / \sum_{i=1}^{c^k} \mu_i \quad (20)$$

Com isso, os ajustes locais contribuem com a inclinação da saída y^k ponderada pelo respectivo grau de ativação. Quanto maior μ_i , $i = 1, \dots, c^k$, maior será a influência da regra i na variação de y^k na direção de cada x_j , $j = 1, \dots, p$.

4.3 Ajuste Recursivo dos Parâmetros

Diferentemente de (Lima et al., 2006a), onde o ajuste dos parâmetros da função linear do conseqüente foi feito globalmente, nesse artigo será utilizado também um ajuste recursivo, de forma

que apenas a última informação coletada é utilizada para a atualização.

Para cada vetor de observações $x \in [0, 1]^p$, a saída y_i da regra $i = 1, 2, \dots, c^k$ pode ser representada pela seguinte igualdade linear:

$$X\gamma = Y \quad (21)$$

onde $X = [x] \in [0, 1]^{m \times p+1}$, $\gamma = [\gamma_i] \in \mathbb{R}^{p+1 \times 1}$ e $Y = [y_i] \in \mathbb{R}^{m \times 1}$.

Considerando o funcional $J_k = J(\gamma^k) = \|X\gamma^k - Y\|_2^2$, sendo γ^k os parâmetros no instante $k = 1, 2, \dots$, a atualização recursiva é definida por uma relação entre J_{k+1} e J_k a partir das informações γ^k, x^{k+1} e y^{k+1} .

Sem perda de generalidade, suprimindo alguns índices por questão de simplicidade e clareza, considere x e y a observação mais recente e a última saída estimada, respectivamente. Dessa forma, definindo as matrizes:

$$X^{k+1} = \begin{pmatrix} X^k \\ x \end{pmatrix}, \quad Y^{k+1} = \begin{pmatrix} Y^k \\ y \end{pmatrix} \quad (22)$$

tem-se a solução que minimiza o funcional J_k no passo k pode ser escrita como (Young, 1984):

$$\gamma^k = P_k b_k \quad (23)$$

onde $P_k = [(X^k)^t X^k]^{-1}$ e $b_k = (X^k)^t Y^k$.

Logo, no passo seguinte $k + 1$ essa expressão é dada por:

$$\gamma^{k+1} = [(X^{k+1})^t X^{k+1}]^{-1} [(X^{k+1})^t Y^{k+1}] \quad (24)$$

Após alguns arranjos algébricos, a equação (24) pode ser escrita em função de P_k, b_k, x e y , da seguinte forma:

$$\gamma^{k+1} = \left[(X^k)^t X^k + x^t x \right]^{-1} \left[(X^k)^t Y^k + x^t y \right] \quad (25)$$

A partir de (23) tem-se que:

$$P_{k+1}^{-1} = P_k^{-1} + x^t x \quad \text{e} \quad b_{k+1} = b_k + x^t y \quad (26)$$

Dessa forma, obtém-se a seguinte igualdade:

$$P_k = P_{k+1} + P_k x^t x P_{k+1} = (I + P_k x^t x) P_{k+1}$$

e, multiplicando por b_{k+1} :

$$P_k b_{k+1} = P_k (b_k + x^t y) = (I + P_k x^t x) P_{k+1} b_{k+1}$$

Finalmente, utilizando (23) e (25):

$$\gamma^{k+1} = (I + P_k x^t x)^{-1} (\gamma^k + P_k x^t y) \quad (27)$$

Logo, dada a atualização no instante k , a expressão acima representa o parâmetro no instante $k + 1$ a partir da última saída y e da observação mais recente x .

5 Aplicação dos Modelos Evolutivos para Previsão de Carga

A previsão de carga é de suma importância para um bom planejamento elétrico e energético. A programação de geração e de intervenções no sistema, e a execução da operação em tempo real são processos cuja qualidade é consequência da precisão dos valores de carga esperados.

Com isso, os algoritmos APE e eTS foram utilizados para previsão de carga a curto prazo de uma empresa de energia elétrica brasileira. Os dados de carga usados são horários e expressos em kilowatts (kW), correspondentes ao mês de Agosto de 2000.

Para a previsão da carga uma hora à frente foram utilizadas 3 horas anteriores, isto é, $y_{t-3}, y_{t-2}, y_{t-1}$, para prever a carga da hora t , representada pela saída y_t . Os modelos eTS e APE foram avaliados para a previsão de 382 horas do mês de Agosto de 2000, isto é, aproximadamente duas semanas, compreendidas entre 0:00h do dia 9 até 0:00h do dia 25. No caso do APE com ajuste global, foram ainda utilizados como dados de treinamento uma semana anterior, ou seja, 1:00h do dia 1^o até 23:00h do dia 8. Em todos os casos, os dados foram normalizados no intervalo $[0, 1]$.

O desempenho dos modelos foi analisado de acordo com 6 critérios: raiz do erro quadrático médio (REQM), erro absoluto médio (EAM), erro relativo médio (ERM) e erro relativo máximo (ERmax), coeficiente de correlação ρ e número de regras ajustadas.

Os resultados, mostrados na tabela 1 e nas figuras 1, 2 e 3 mostraram que os modelos apresentaram desempenhos similares. Entretanto, o APE teve um menor número de regras ajustadas, em ambas versões.

Tabela 1: Desempenho dos modelos.

Desempenho	Modelos		
	APE	APEr*	eTS
REQM($10^5 kW$)	3,0988	3,3425	4,3447
EAM($10^5 kW$)	2,1710	2,2967	2,3537
ERM(%)	4,5922	4,7782	4,7278
ERmax(%)	29,2213	64,9680	82,3749
ρ	0,9685	0,9625	0,9360
Número de regras	4	5	7

* APE com atualização recursiva do consequente

6 Conclusão

Esse artigo descreveu uma modelagem alternativa baseada na noção de aprendizado participativo. O método inclui um algoritmo de agrupamento não supervisionado, capaz de identificar dinamicamente o número de grupos, e um mecanismo que atualiza em tempo real a estrutura das regras. Além disso, como avanços à idéia proposta

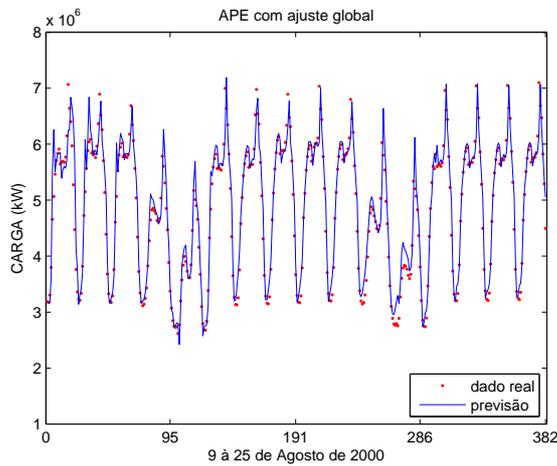


Figura 1: APE com ajuste global.

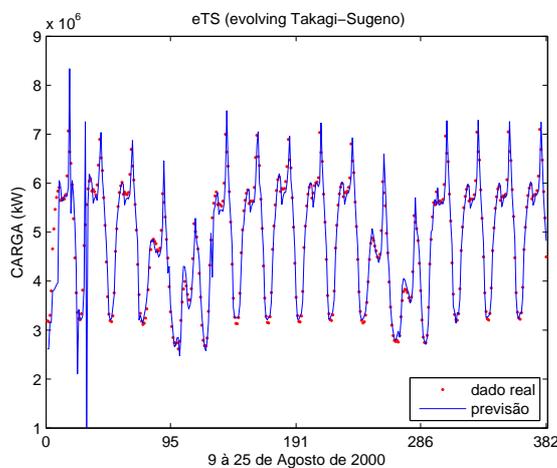


Figura 2: Algoritmo eTS

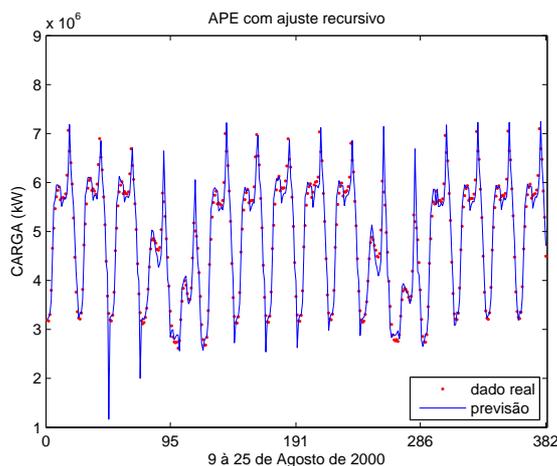


Figura 3: APE com ajuste recursivo.

originalmente, foram mostrados alguns aspectos do comportamento do algoritmo bem como implementado o ajuste recursivo dos consequentes.

O desempenho do APE, tanto na versão com ajuste global quanto recursivo, foi comparado ao

algoritmo eTS, no qual baseia-se a estrutura de regras do APE. As comparações foram feitas na previsão de carga horária de uma empresa de energia elétrica brasileira. Os resultados mostraram que o APE proporciona um modelo eficiente, leve e parcimonioso na identificação e previsão de padrões contidos na série.

Agradecimentos

Oa autores agradecem o suporte da Fundação de Amparo Pesquisa do Estado de São Paulo (FAPESP) e do Conselho Nacional de Pesquisa e Desenvolvimento (CNPq).

Referências

- Anderson, K. L., Blackenship, G. L. and Lebow, L. G. (1988). A rule-based adaptive pid controller, *Proc. 27th IEEE Conf. Decision Control*, New York, USA, pp. 564–569.
- Angelov, P. P. (2002). *Evolving Rule-Based Models: A Tool for Design of Flexible Adaptive Systems*, ISBN 3-7908-1457-1, Springer-Verlag, Heidelberg, New York.
- Angelov, P. P. and Filev, D. (2004). An approach to on-line identification of evolving Takagi-Sugeno fuzzy models, *IEEE Trans. on Systems, Man and Cybernetics, part B* **34**(1): 484–498.
- Angelov, P. P. and Guthke, R. (1997). A GA-based approach to optimization of bioprocesses described by fuzzy rules, *Journal Bioprocess Engineering* **16**: 299–301.
- Chiu, S. L. (1994). Fuzzy model identification based on cluster estimation, *Journal of Intelligence and Fuzzy System* **2**: 267–278.
- Jang, J. S. R. (1993). Anfis: Adaptive network-based fuzzy inference systems, *IEEE Trans. on Systems, Man and Cybernetics* **23**: 665–685.
- Lima, E., Ballini, R. and Gomide, F. (2006a). Modelagem evolucionária participativa nebulosa, *XVI Congresso Brasileiro de Automática* pp. 2619–2624.
- Lima, E., Ballini, R. and Gomide, F. (2006b). Participatory evolving fuzzy modeling, *International Symposium on Evolving Fuzzy Systems* pp. 36–41.
- Takagi, T. and Sugeno, M. (1985). Fuzzy identification of systems and its application to modeling and control, *IEEE Trans. on Systems, Man and Cybernetics* **15**: 116–132.
- Yager, R. R. (1990). A model of participatory learning, *IEEE Trans. on Systems, Man and Cybernetics* **20**: 1229–1234.
- Yager, R. R. (2004). Participatory learning: A paradigm for more human like learning, *The 2004 IEEE International Conference on Fuzzy Systems FUZZ-IEEE* **1**: 79–84.
- Young, P. C. (1984). *Recursive Estimation and Time-Series Analysis*, Springer-Verlag, Berlin.