

Um Método de Particionamento Baseado no *K-Means* para Preditores Probabilísticos *Fuzzy*

Marcelo Andrade Teixeira^{1,2}

Gerson Zaverucha¹

¹: Programa de Engenharia de Sistemas e Computação - COPPE
Universidade Federal do Rio de Janeiro (UFRJ) - {mat, gerson}@cos.ufrj.br
Caixa Postal 68511, CEP 21945-970, Rio de Janeiro, Brasil

²: Centro de Pesquisas de Energia Elétrica (CEPEL) - mat@cepel.br
Caixa Postal 68007, CEP 21944-970, Rio de Janeiro, Brasil

Resumo

Nós apresentamos uma modificação do método de particionamento conhecido como K-Means (KM) para determinar regiões fuzzy em Preditores Probabilísticos Fuzzy (PPF's). PPF's são modificações de classificadores probabilísticos discretos, como o Naive Bayes Classifier e o Hidden Markov Model, a fim de habilitá-los a prever valores contínuos. Vários PPF's utilizando esse particionamento baseado no KM são aplicados à tarefa de previsão de séries de carga elétrica mensal e comparados com sucesso a dois modelos de filtro de Kalman e dois métodos de previsão tradicionais, Box-Jenkins e amortecimento exponencial de Winters. As séries temporais empregadas apresentam um comportamento de mudança abrupta e significativa em seus últimos anos, como quando ocorre um racionamento de energia.

1. Introdução

Uma possível abordagem para se prever uma série temporal ou, de modo mais geral, qualquer valor contínuo (regressão) é através da estimação da função de densidade de probabilidade (fdp) da variável aleatória contínua que se deseja prever. Uma vez de posse dessa função de densidade então temos que a previsão é dada pelo valor esperado da variável aleatória contínua (o que corresponde a calcular a média da distribuição).

A estimação da densidade, ou seja, o ajuste de uma função de densidade aos dados [16], pode ser paramétrica ou não-paramétrica. No caso do ajuste paramétrico, é necessária a especificação da forma da

densidade (por exemplo, Gaussiana) e a estimação de seus parâmetros. Já o ajuste não-paramétrico (por exemplo, histogramas [16]) fornece um algoritmo consistente para quase qualquer densidade contínua e evita o passo de especificação. Se considerarmos o histograma para o ajuste não-paramétrico, vemos que a estimação da densidade contínua é feita pela discretização dos dados: para cada valor contínuo existe um correspondente valor discreto representando o intervalo que contém o valor contínuo.

A metodologia paramétrica para previsão através do valor esperado foi amplamente usada, por exemplo, em modelos de filtro de Kalman [9], [25]. Não existem muitos trabalhos na literatura que usam a metodologia não-paramétrica para previsão através do valor esperado, mas a pesquisa neste campo produziu alguns sistemas interessantes: *Naive Bayes for Regression* (NBR) [5], *Markov Model for Regression* (MMR) [17], [19], *Hidden Markov Model for Regression* (HMMR) [19] e *Multi-Hidden Markov Model for Regression* (MHMMR) [21]. Esses Preditores Probabilísticos Discretos (PPD's) usam discretização e Redes Bayesianas (RB's) [15] ou Redes Bayesianas Dinâmicas (RBD's) [7], [14], [15] com apenas variáveis discretas (por exemplo, *Naive Bayes Classifier* [4] e *Hidden Markov Model* [13]) para estimar fdp's e então fazer previsões contínuas.

Em nossos trabalhos anteriores [18], [19], [21], estendemos esses PPD's empregando a fuzzificação [11], [24] como uma generalização da discretização. A fuzzificação dos dados contínuos funciona da seguinte forma: para cada valor contínuo v existe um correspondente conjunto de valores *fuzzy* $\{r \mid m_r(v) > 0\}$, onde $m_r(v)$ é uma função de pertinência que retorna

um número real em $[0, 1]$ para uma região *fuzzy* r e um valor contínuo v . Assim obtivemos vários Preditores Probabilísticos *Fuzzy* (PPF's) que seguem essa idéia: *Fuzzy Bayes Predictor* (FBP) [18], *Fuzzy Markov Predictor* (FMP) [18], *Fuzzy Hidden Markov Predictor* (FHMP) [19] e *Fuzzy Multi-Hidden Markov Predictor* (FMHMP) [21]. Em PPF's, o ajuste não-paramétrico é feito por RB's ou RBD's utilizando variáveis e probabilidades *fuzzy* [22], [26].

Tanto os PPD's quanto os PPF's partem do princípio de que o espaço dos dados contínuos foi previamente particionado em intervalos ou regiões *fuzzy*. Em ambos os casos, a abordagem mais simples é efetuar um particionamento uniforme do espaço contínuo: na discretização, o espaço é dividido em intervalos de mesmo tamanho; na fuzzificação, regiões *fuzzy* com o mesmo formato são distribuídas uniformemente no espaço. *Density Trees* (DT's) [20], [23], *K-Means* (KM) [8] e suas respectivas versões *fuzzy* são métodos de particionamento que não possuem essa restrição de uniformidade.

O artigo é organizado como segue. Na seção 2 são revistos os particionamentos discreto e *fuzzy* usando DT's [20]. Na seção 3 é revisado o particionamento discreto usando KM e é apresentada sua versão *fuzzy*. Na seção 4, PPD's e PPF's usando diferentes tipos de particionamento são aplicados à tarefa de predição de carga elétrica mensal e comparados a dois modelos de filtro de Kalman, STAMP [9] e BATS [25], e dois métodos de predição tradicionais, Box-Jenkins [2] e amortecimento exponencial de Winters [12]. Finalmente, na seção 5 são discutidas algumas conclusões e trabalho futuro.

2. Particionamento *fuzzy* através de Density Trees

DT's fazem um particionamento recursivo do espaço contínuo com o propósito de estimar (de forma não-paramétrica) a fdp "f" de uma variável aleatória contínua. Em [23], uma DT é utilizada para dividir (discretizar) o espaço contínuo em intervalos da seguinte maneira:

1. comece com N valores (exemplos) contínuos em um intervalo (o nó raiz) que cobre o domínio inteiro de "f";
2. divida o intervalo corrente em dois intervalos (filhos) de mesmo tamanho se o corrente possui ao menos \sqrt{N} valores e sua distância ao nó raiz não excede $\lfloor (\log_2 N)/4 \rfloor$ (a distância entre um nó/intervalo e seus nós/intervalos filhos é igual a 1);

3. repita este processo para cada novo intervalo enquanto as condições forem satisfeitas.

A figura 1 mostra um exemplo do particionamento feito por uma DT. Neste exemplo, os intervalos finais obtidos são $[0.0; 0.25]$, $[0.25; 0.5]$ e $[0.5; 1.0]$.

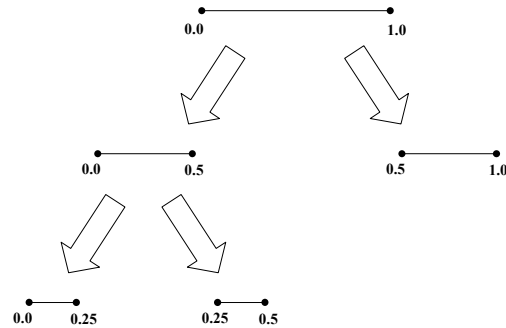


Figura 1: Particionamento feito por uma *Density Tree*

A versão *fuzzy* desse algoritmo [20] para a obtenção de uma DT funciona da mesma maneira que o original na divisão de intervalos. A única diferença é que são colocadas regiões *fuzzy* nesses intervalos:

1. o intervalo/nó raiz contém duas regiões *fuzzy* triangulares cujos máximos e mínimos (valores de pertinência iguais a 1 e 0, respectivamente) são os limites do intervalo, e a interseção das regiões *fuzzy* é o centro do intervalo;
2. cada intervalo corrente contém parte de duas regiões *fuzzy* triangulares cujos máximos e mínimos são os limites do intervalo, e a interseção dessas regiões *fuzzy* é o centro do intervalo;
3. quando o intervalo corrente é dividido em dois intervalos de mesmo tamanho (se as condições do passo 2 da DT são satisfeitas), os mínimos das duas regiões *fuzzy* contidos no intervalo corrente são mudados para o centro desse intervalo, e uma nova região *fuzzy* triangular é inserida com seu máximo igual ao centro do intervalo corrente e mínimos iguais aos limites desse intervalo;
4. este processo é repetido para cada novo intervalo enquanto as condições forem satisfeitas.

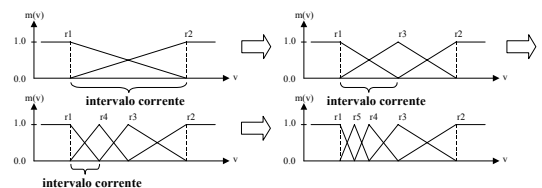


Figura 2: Particionamento feito pela versão *fuzzy* do algoritmo para DT

Um exemplo da criação de regiões *fuzzy* através desse método de particionamento é mostrado na figura 2. Partindo de duas regiões *fuzzy*, r_1 e r_2 , o algoritmo adiciona as novas regiões *fuzzy* r_3 , r_4 e r_5 nesta ordem.

Na figura 3 é visto o mesmo exemplo anteriormente apresentado para uma DT, mas agora usando sua versão *fuzzy*. As regiões *fuzzy* finais obtidas são r_1 , r_2 , r_3 e r_4 .

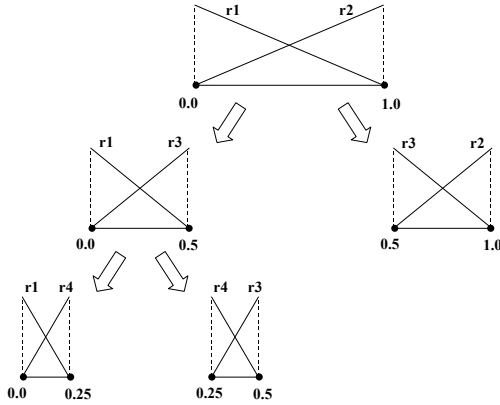


Figura 3: Particionamento *fuzzy* usando uma DT

3. Versão *fuzzy* do K-Means

O método de particionamento conhecido como *K-Means* (KM) tem por objetivo construir k intervalos de forma a minimizar a soma das distâncias $d_{i,j}$ ($d_{i,j} = |x_i - c_j|$) de cada valor (exemplo) contínuo x_i contido em um intervalo para o centro de gravidade c_j do intervalo:

1. inicialize o centro de gravidade c_j de cada intervalo j (por exemplo, pode-se escolher os intervalos de forma que cada um tenha a mesma quantidade de exemplos);
2. calcule (para $i = 1, \dots, N$; $j = 1, \dots, k$) $m_{i,j} = \{1 \text{ se } d_{i,j} < d_{i,r} \text{ para todo } r \neq j; 0 \text{ caso contrário}\}$, ou seja, a pertinência $m_{i,j}$ de um valor x_i em um intervalo j ;
3. calcule o centro de gravidade c_j de cada intervalo j : $c_j = (\sum_{i=1}^N m_{i,j} \cdot x_i) / (\sum_{i=1}^N m_{i,j})$;
4. repita os passos 2-3 até não haver alterações nos centros de gravidade (ou atingir um número máximo de iterações).

O ponto de interseção entre intervalos adjacentes j e $j+1$ é dado por $(c_j + c_{j+1}) / 2$. Supondo que a e b sejam os limites inferior e superior do domínio contínuo, respectivamente, os intervalos obtidos são $[a; (c_1 + c_2) / 2]$, $[(c_1 + c_2) / 2; (c_2 + c_3) / 2]$, ... e $[(c_{k-1} + c_k) / 2; b]$.

Uma outra possibilidade para inicialização dos centros de gravidade dos intervalos é usar o particionamento feito por uma DT. Se os k intervalos obtidos pela DT forem $[l_1; l_2]$, $[l_2; l_3]$, ... e $[l_k; l_{k+1}]$

então os centros de gravidade iniciais podem ser $c_1 = (l_1 + l_2) / 2$, $c_2 = (2 \cdot l_2 - c_1)$, $c_3 = (2 \cdot l_3 - c_2)$, ... e $c_k = (2 \cdot l_k - c_{k-1})$. Com esses centros consegue-se manter os mesmos pontos de interseção entre intervalos adjacentes produzidos pela DT pois $l_{j+1} = (c_j + c_{j+1}) / 2$.

A versão *fuzzy* do KM apresentada aqui utiliza k regiões *fuzzy* triangulares ao invés de intervalos. O máximo (valor de pertinência igual a 1) de uma região *fuzzy* j é representado por c_j . Dado duas regiões *fuzzy* adjacentes j e $j+1$, um mínimo (valor de pertinência igual a 0) da região *fuzzy* j é igual a c_{j+1} , um mínimo da região *fuzzy* $j+1$ é igual a c_j , e o ponto com valor de pertinência igual a 0.5 (para ambas as regiões) é igual a $(c_j + c_{j+1}) / 2$. O algoritmo consiste nos seguintes passos:

1. inicialize o máximo c_j de cada região *fuzzy* j (por exemplo, pode-se escolher esses máximos de forma que sejam iguais aos centros de gravidade da inicialização do KM onde cada intervalo tem a mesma quantidade de exemplos);
2. calcule (para $i = 1, \dots, N$; $j = 1, \dots, k$) $m_{i,j} = m_j(x_i)$, onde $m_j(\cdot)$ é a função de pertinência da região *fuzzy* j (que depende dos máximos de j e das regiões adjacentes a j);
3. calcule o máximo c_j de cada região *fuzzy* j : $c_j = (\sum_{i=1}^N m_{i,j} \cdot x_i) / (\sum_{i=1}^N m_{i,j})$;
4. repita os passos 2-3 até não haver alterações nos máximos (ou atingir um número máximo de iterações).

Alguns aspectos devem ser ressaltados sobre esse algoritmo:

- embora não mencionado explicitamente, os máximos das regiões *fuzzy* cujos valores são os limites inferior e superior do domínio contínuo não devem ser alterados pelo algoritmo a fim de preservar a informação sobre os limites do domínio;
- de forma semelhante ao que acontece no KM ao utilizar um DT no passo de inicialização, o particionamento feito pela versão *fuzzy* do algoritmo para DT pode ser usado para inicializar os máximos das regiões *fuzzy*.

A versão *fuzzy* do algoritmo KM desenvolvida neste trabalho é diferente do *Fuzzy K-Means* (FKM) que aparece na literatura [1], [3]. No FKM, a pertinência $m_{i,j}$ presente no algoritmo do KM é dada por

$$m_{i,j} = d_{i,j}^{2/(\phi-1)} / (\sum_{r=1}^k d_{i,r}^{2/(\phi-1)})$$

onde ϕ ($1 < \phi < \infty$) é o grau de sobreposição entre os k grupos de dados. Como o FKM não usa regiões *fuzzy* triangulares, ele não poderia utilizar a versão *fuzzy* do algoritmo para DT no passo de inicialização. Por isso

optamos pelo uso de uma versão *fuzzy* do KM diferente do FKM.

4. Resultados experimentais

PPD's e PPF's foram aplicados à tarefa de predição de 1-passo-adiante de séries de carga elétrica mensal e foram comparados a dois modelos de filtro de Kalman, STAMP e BATS, e dois métodos de predição tradicionais, Box-Jenkins e amortecimento exponencial de Winters.

Foram empregadas três séries (figuras 4, 5 e 6) de carga elétrica mensal (5 x 12 meses de dados para treinamento e 3 x 12 meses para teste). Essas séries foram obtidas de empresas brasileiras de energia elétrica e apresentam um comportamento de mudança abrupta e significativa em seus últimos anos, como quando ocorre um racionamento de energia.

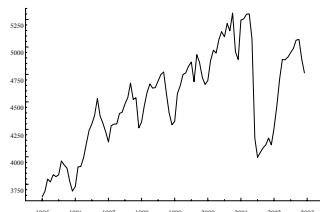


Figura 4: Série 1 de carga elétrica mensal

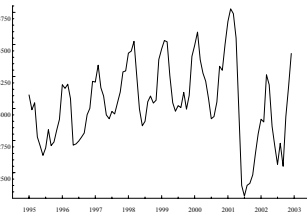


Figura 5: Série 2 de carga elétrica mensal

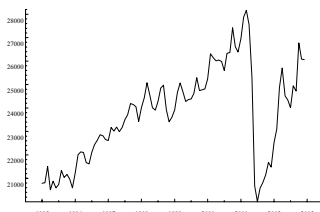


Figura 6: Série 3 de carga elétrica mensal

A métrica de erro utilizada foi a MAPE (Mean Absolute Percentage Deviation):

$$MAPE = \left(\sum_{i=1}^n |e_i| \right) / n$$

onde,

$$e_i = ((desejado_i - previsto_i) / desejado_i) * 100\%$$

n = número de exemplos de teste.

Todos os sistemas (PPD's, PPF's, modelos de filtro de Kalman e métodos tradicionais) fizeram uso dos últimos 3 anos das séries como o conjunto de teste, e os 5 anos anteriores como o conjunto de treinamento para efetuar a predição de 1-passo-adiante do próximo mês que corresponde ao primeiro mês do conjunto de teste. Para se prever cada mês do conjunto de teste os sistemas são retreinados com os 5 anos que precedem o mês sendo previsto. Esse retreinamento é feito devido ao comportamento de mudança abrupta nas séries.

Quatro casos foram considerados para a distribuição dos intervalos ou regiões *fuzzy* (triangulares) no espaço contínuo de um PPD ou PPF: distribuição uniforme das partições (intervalos ou regiões *fuzzy*); partições distribuídas de acordo com o algoritmo para DT; partições obtidas pelo KM; inicialização das partições através do algoritmo para DT seguido pelo ajuste das mesmas pelo KM. Os erros de predição para cada uma dessas possibilidades são mostrados nas figuras 7, 8, 9 e 10. Nessas figuras, "F#HMP" ou "#HMMP" indicam um FMHMP ou um MHMMP com o número de variáveis de estado igual a "#". Em cada uma dessas figuras os erros são referentes às predições dos três anos do conjunto de teste (para as três séries) e sempre são comparados PPD's e PPF's com modelos de filtro de Kalman e métodos tradicionais.

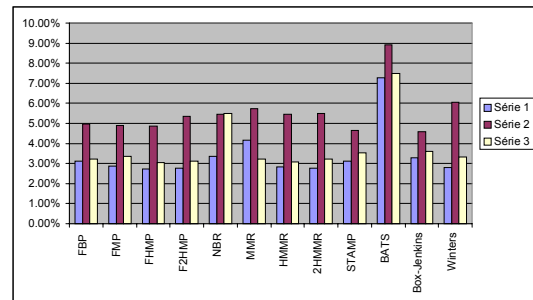


Figura 7: MAPE de PPD's e PPF's com distribuição uniforme das partições

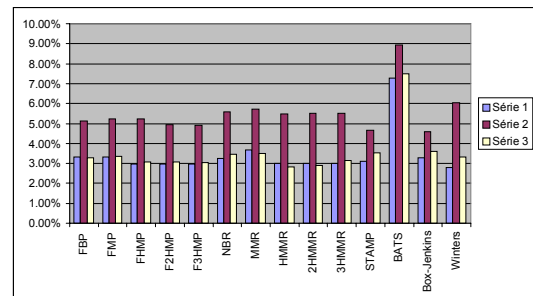


Figura 8: MAPE de PPD's e PPF's com partições distribuídas através de DT

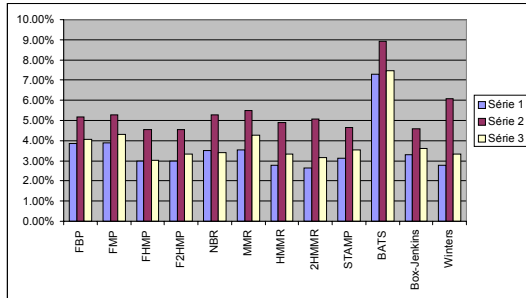


Figura 9: MAPE de PPD's e PPF's com partições obtidas pelo KM

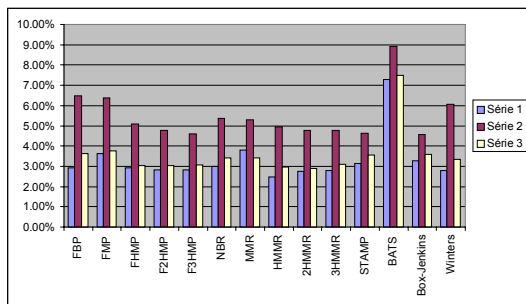


Figura 10: MAPE de PPD's e PPF's com partições obtidas por DT e KM

Forward validation (FV) [10] foi utilizado para a seleção do número de atributos (número de componentes da observação) e o número de partições nos PPD's e PPF's, exceto quando a escolha do número de partições é feita por um algoritmo para DT.

As três séries foram diferenciadas para os PPD's e PPF's, ou seja, cada série temporal diferenciada é obtida por $Y_t = Z_t - Z_{t-1}$, onde Z_t corresponde a um valor de cada uma das séries originais. Para os PPD's e PPF's baseados em RBD's, os tempos $t-1$ e t representam o mesmo mês em anos consecutivos.

Para Box-Jenkins, a identificação da diferenciação usada no modelo ARIMA é feita por um algoritmo baseado no método Augmented Dickey-Fuller [6]. A identificação do modelo para os dados diferenciados é realizada pela minimização do BIC (*Bayesian Information Criterion*) sobre um número de modelos ARIMA estimados aproximadamente.

As figuras 7, 8, 9 e 10 apresentam os erros de predição de 1-passo-adiante para PPD's e PPF's com a escolha do número de entradas feita pelo FV. Comparando-se cada preditor que usa a abordagem uniforme com o mesmo preditor que usa DT (totalizando 8 preditores), a abordagem uniforme obteve 13 vitórias (de 24 possibilidades = 8 preditores x 3 séries), onde cada uma dessas vitórias representa a constatação de que um preditor específico usando a abordagem uniforme obteve um erro de predição

menor que o mesmo preditor usando DT para alguma série. Nessa mesma comparação, a abordagem por DT obteve 7 vitórias e ocorreram 4 empates. A princípio isso pode sugerir que a abordagem uniforme seja mais adequada que a de DT, mas também é preciso considerar que o uso de DT é computacionalmente mais rápido que o processo de escolha do número de partições uniformes. Por exemplo, o tempo de computação para o uso de PPD's e PPF's com três variáveis de estado para a abordagem uniforme é extremamente maior do que a abordagem por DT.

Já na comparação da abordagem uniforme com a que usa KM (totalizando 8 preditores) evidencia-se os seguintes resultados: para a uniforme 12 vitórias e a do KM 12 vitórias, o que nos leva a crer que o uso de KM não necessariamente melhora nossas predições. Comparando-se a abordagem que usa DT com aquela que combina DT e KM (totalizando 10 preditores), ocorrem 9 vitórias para a de DT e 21 vitórias para a de DT-KM (de 30 possibilidades = 10 preditores x 3 séries), o que sugere que o KM pode melhorar o particionamento feito pela DT. Comparando-se DT-KM com uniforme (total de 8 preditores) temos 14 vitórias para o primeiro e 10 vitórias para o segundo. Considerando DT-KM versus KM (total de 8 preditores) temos 14 vitórias para o primeiro, 9 vitórias para o segundo e 1 empate. Pode-se concluir que o uso conjunto de DT com KM é uma abordagem eficiente e rápida em PPD's e PPF's de estruturas complexas: por exemplo, o F3HMP e o 3HMMR obtiveram resultados satisfatórios quando comparados aos modelos de filtro de Kalman e métodos tradicionais.

7. Conclusão e trabalho futuro

Nós apresentamos uma modificação do método de particionamento conhecido como *K-Means* (KM) para determinar regiões *fuzzy* em Preditores Probabilísticos *Fuzzy* (PPF's). Preditores Probabilísticos Discretos (PPD's) e PPF's usando suas respectivas versões de KM foram aplicados à tarefa de predição de carga elétrica mensal onde as séries temporais empregadas apresentam um comportamento de mudança abrupta e significativa em seus últimos anos, como quando ocorre um racionamento de energia. O uso conjunto do KM com o particionamento feito por Density Trees (DT's) também foi testado (tanto para o caso discreto quanto o *fuzzy*).

De forma geral, os PPD's e PPF's obtiveram resultados competitivos quando comparados com os dois modelos de filtro de Kalman, STAMP e BATS, e os dois métodos tradicionais para predição, Box-Jenkins e amortecimento exponencial de Winters.

Analisando as figuras 7, 8, 9 e 10 podemos ver que o uso conjunto de DT com KM é uma abordagem eficiente e rápida em PPD's e PPF's de estruturas complexas.

Como trabalho futuro, gostaríamos de testar outras séries temporais e automatizar a escolha de qual PPD ou PPF é o mais adequado para uma determinada série.

8. Referências

- [1] J. C. Bezdek. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York, 1981.
- [2] G.E.P. Box, G.M. Jenkins and G.C. Reinsel, Time Series Analysis: Forecasting & Control, Prentice Hall, 1994.
- [3] J. J De Gruijter and B. McBratney. A modified fuzzy k means for predictive classification. In: Bock,H.H.(ed) Classification and Related Methods of Data Analysis. pp. 97-104. Elsevier Science, Amsterdam, 1988.
- [4] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," Machine Learning 29(2/3):, pp. 103-130, 1997.
- [5] E. Frank, L. Trigg, G. Holmes and I.H. Witten, "Naive Bayes for regression," Machine Learning, Vol.41, No.1, pp.5-25, 1999.
- [6] W.A. Fuller, Introduction to statistical time series, John Wiley, 1996.
- [7] Z. Ghahramani, "Learning dynamic Bayesian networks," in Adaptive Processing of Sequences and Data Structures, Lecture Notes in Artificial Intelligence, C.L. Giles and M. Gori (eds.), Berlin, Springer-Verlag, 1998, pp. 168-197.
- [8] J. Hartigan and Wong. A K-Means clustering algorithm, ALGORITHM AS 136. Applied Statistics, Vol. 28, Number 1, pp. 100-108, 1979.
- [9] A.C. Harvey, Forecasting, structural time series models and the Kalman filter, Cambridge University Press, 1994.
- [10] J.S.U. Hjorth. Computer Intensive Statistical Methods. Validation Model Selection and Bootstrap. Chapman & Hall. 1994.
- [11] J.M. Mendel, "Fuzzy logic systems for engineering: a tutorial," Proceedings of the IEEE, vol.83, pp.345-377, 1995.
- [12] D.C. Montgomery, L.A. Johnson and J.S. Gardiner, Forecasting and Time Series Analysis, McGraw-Hill Companies, 1990.
- [13] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proc. of the IEEE, vol. 77, no. 2, pp. 257-286, 1989.
- [14] S. Roweis and Z. Ghahramani, "A unifying review of linear Gaussian models," Neural Computation 11(2), pp. 305-345, 1999.
- [15] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, Prentice Hall, 2nd edition, 2002.
- [16] D.W. Scott, "Density Estimation", in P. Armitage & T. Colton, editors, Encyclopedia of Biostatistics, pp. 1134-1139. J. Wiley & Sons, Chichester, 1998.
- [17] M.A. Teixeira, K. Revoredo and G. Zaverucha, "Hidden Markov model for regression in electric load forecasting," ICANN/ICONIP, Istanbul, Turkey, June 26-29, pp. 374-377. 2003.
- [18] M.A. Teixeira and G. Zaverucha, "Fuzzy Bayes and Fuzzy Markov Predictors" Journal of Intelligent and Fuzzy Systems. IOS Press, Amsterdam, The Netherlands, volume 13, numbers 2-4, pp. 155-165, 2003.
- [19] M.A. Teixeira and G. Zaverucha, "Fuzzy hidden Markov predictor in electric load forecasting," International Joint Conference on Neural Networks, Vol. 1, pp. 315-320, 2004.
- [20] M.A. Teixeira and G. Zaverucha, "A Partitioning Method for Fuzzy Probabilistic Predictors," ICONIP, Lecture Notes in Computer Science Vol. 3316, pp. 929-934, Springer, 2004.
- [21] M.A. Teixeira and G. Zaverucha, "Fuzzy multi-hidden Markov predictor in electric load forecasting," International Joint Conference on Neural Networks, accepted, 2005.
- [22] T. Terano, K. Asai and M. Sugeno, Fuzzy Systems Theory and Its Applications, Academic Press, Incorporated, 1992.
- [23] S. Thrun, J. Langford, and D. Fox. "Monte Carlo hidden Markov models: Learning non-parametric models of partially observable stochastic processes". Proc. of the International Conference on Machine Learning (ICML), pp. 415-424, 1999.
- [24] L. Wang and J.M. Mendel, "Generating fuzzy rules by learning from examples," IEEE Transactions on Systems, Man and Cybernetics, v.22, n.6, 1992.
- [25] M. West and J. Harrison, Bayesian Forecasting and Dynamic Models (2nd ed.), Springer, 1997.
- [26] L.A. Zadeh, "Probability measures of fuzzy events," Jour. Math. Analysis and Appl. 23, pp. 421-427, 1968.