

# Um Estudo das Funções de Base Radial aplicadas à Previsão de Cargas Elétricas via Redes Neurais Artificiais

Alessandra B. Altran, Mara L. M. Lopes, Carlos R. Minussi

Depto de Engenharia Elétrica, FEIS, UNESP,  
15385-000, Ilha Solteira, SP

[lealtran@yahoo.com.br](mailto:lealtran@yahoo.com.br), [mara@aluno.feis.unesp.br](mailto:mara@aluno.feis.unesp.br), [minussi@dee.feis.unesp.br](mailto:minussi@dee.feis.unesp.br)

Francisco Villarreal

Depto de Matemática, FEIS, UNESP,  
15385-000, Ilha Solteira, SP  
[villa@fqm.feis.unesp.br](mailto:villa@fqm.feis.unesp.br)

**Resumo:** *Este trabalho tem por objetivo o desenvolvimento de uma metodologia para o treinamento de redes neurais artificiais feedforward baseado no algoritmo Retropropagação utilizando diferentes funções de ativação aplicadas ao problema de previsão de cargas elétricas a curto-prazo. O algoritmo Retropropagação é considerado na literatura um benchmark em termos de precisão. Vem se tornando uma das principais ferramentas utilizadas para resolver problemas reais, por exemplo, previsão de cargas elétricas. Deste modo, neste trabalho será efetuada uma análise na arquitetura do algoritmo Retropropagação no qual será investigada a aplicação de diferentes funções de ativação à rede neural artificial de forma a identificar como essas funções atuam na saída da rede neural artificial. Como forma de ilustração da metodologia proposta, são apresentados os resultados (análise comparativa) da previsão de cargas elétricas, considerando-se os dados históricos de uma companhia do setor elétrico brasileiro.*

**Palavras-chave:** Algoritmo Retropropagação, Funções de Base Radial, Previsão de Cargas Elétricas, Redes Neurais Artificiais.

## Introdução

O conhecimento futuro do perfil das cargas elétricas é fator essencial no planejamento e operação de sistemas elétricos de potência. Estas atividades dependem efetivamente do comportamento da carga elétrica. Estas podem ser previstas através de procedimentos estatísticos usados para a construção de séries temporais baseadas em dados históricos armazenados nas bases de dados das empresas do setor elétrico. Na literatura encontram-se técnicas para a previsão de carga elétrica realizadas através de métodos estatísticos tradicionais, como os métodos ARIMA de Box & Jenkins [2], [3]. Atualmente, tem sido intensificado o uso de abordagens alternativas para a previsão de cargas, baseadas no conceito de inteligência artificial [1], em destaque, as Redes Neurais Artificiais (RNA), cuja vantagem está no fato

de que o conhecimento é extraído da base de dados sem a necessidade do conhecimento prévio do modelo das cargas. Este conhecimento é extraído em função de estímulos (entrada/saída), adequadamente aplicados à rede neural, constituindo a fase de treinamento.

As redes neurais artificiais são concepções em *hardware* e/ou *software* que exibem capacidade de aprender com a experiência. Para se conseguir os resultados desejados, ou seja, que a rede apresente condições de efetuar diagnósticos complexos, como é caso do reconhecimento de padrão, previsão, etc; tais redes deverão apresentar configurações formadas por várias unidades de "neurônios" dispostas em fileiras, compondo um arranjo complexo de interligações [5]. As interligações são formadas por pesos (sinapses) que devem ser ajustados em função de um conjunto de padrões que produzam saídas desejadas. Esta atividade é definida como sendo treinamento elaborado *off-line*. Uma vez ajustada a rede deverá ser capaz de realizar, para padrões não constantes no conjunto de treinamento, um diagnóstico com precisão satisfatória. Este diagnóstico pode ser efetuado sem custo computacional. Isto, a princípio, pode ser visto como uma possibilidade de viabilização da análise em tempo real.

As redes neurais *feedforward* utilizando o algoritmo retropropagação (*backpropagation*) [18], é uma das metodologias mais empregadas na literatura. Desta forma, é fundamental investigar como funciona a arquitetura do algoritmo retropropagação quando a ele se aplicam diferentes funções de ativação (funções de base radial e função sigmoidal), ou seja, é importante que seja feita uma análise criteriosa do comportamento da saída da rede neural com relação à precisão dos resultados e ao tempo de processamento relativos às diferentes funções de ativação.

## Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência [6], [15].

Uma rede neural artificial consiste de elementos de

processamento (neurônios) e suas conexões (sinapses), e cada conexão possui um peso que determina sua contribuição na decisão do disparo, controlando dessa forma, os estímulos.

## Estrutura da Rede Neural

Nesta seção será apresentada uma rede neural multicamadas *feedforward*, cujo treinamento será realizado através do algoritmo retropropagação, que é um procedimento sistemático de adaptação de pesos de redes neurais baseado no método do gradiente descendente [19].

O  $i$ -ésimo elemento de saída (neurônio)[19] é uma combinação linear dos elementos de entrada  $x_j$  conectados ao elemento  $i$  através do peso  $w_{ij}$ . Então:

$$J_i = \sum_j w_{ij} x_j \quad (1)$$

Cada elemento possui um termo *bias*  $w_0$  alimentado por uma entrada constante  $x_0 = +1$ . A saída linear  $J_i$  é finalmente convertida numa função não-linear [16], [19].

## Treinamento da Rede Neural

O algoritmo Retropropagação é esquematizado através dos seguintes procedimentos:

1. Apresentação de um padrão  $X$  a rede, o qual fornece uma saída  $Y$ ;
2. Cálculo do erro (diferença entre o valor desejado e a saída) para cada saída;
3. Determinação do erro retropropagado pela rede associado à derivada parcial do erro quadrático de cada elemento relacionado aos pesos;
4. Ajuste dos pesos de cada elemento;
5. Apresenta-se um novo padrão a rede e repete-se o processo até que haja a convergência da rede, ou seja, ( $|\text{erro}| \leq \text{tolerância arbitrada}$ ).

Inicialmente os valores dos pesos são adotados como números aleatórios considerando o intervalo  $\{0,1\}$  [19]. O algoritmo Retropropagação consiste na adaptação dos pesos de tal forma que o erro quadrático da rede seja minimizado. A soma dos erros quadráticos instantâneos de cada neurônio da última camada (saída da rede) é dada por:

$$e^2 = \sum_{i=1}^{no} e_i^2 \quad (2)$$

em que:

- $e_i = d_i - s_i$ ;
- $d_i =$  saída desejada do  $i$ -ésimo elemento da última camada;
- $s_i =$  saída do  $i$ -ésimo elemento da última camada;
- $no =$  número de neurônios da última camada.

Considerando o  $i$ -ésimo neurônio da rede e usando o método do gradiente descendente, o ajuste dos pesos é formulado por [19]:

$$V_i(h+1) = V_i(h) + q_i(h) \quad (3)$$

sendo:

- $q_i(h) = -g[\tilde{N}_i(h)]$ ;
- $g =$  taxa de treinamento;
- $\tilde{N}_i(h) =$  gradiente do erro quadrático relacionado aos pesos do neurônio  $i$ ;
- $V_i \triangleq$  vetor contendo os pesos do neurônio  $i$ ;
- $= [w_{0i} \ w_{1i} \ w_{2i} \ \dots \ w_{ni}]^T$ .

A direção adotada em (3), para a minimização da função objetivo do erro quadrático, corresponde a direção oposta do gradiente. O parâmetro  $g$  determina o comprimento do vetor  $q_i(h)$  e serve como controle de estabilidade do processo iterativo [12].

Desta maneira, obtém-se o seguinte esquema para adaptação dos pesos [13], [19]:

$$V_i(h+1) = V_i(h) + 2g b_i X_i \quad (4)$$

sendo:

- $X_i \triangleq$  vetor padrão;
- $= [x_{0i} \ x_{1i} \ x_{2i} \ \dots \ x_{ni}]^T$ .

Se o  $i$ -ésimo elemento está na última camada, então:

$$b_i = s_i e_i \quad (5)$$

Se o  $i$ -ésimo elemento está nas outras camadas, então:

$$b_i = s_i \sum_{j \in Q(i)} w_{ij} b_j \quad (6)$$

em que:

- $s_i \triangleq$  derivada da função de ativação, relativa a  $J_i$ ;
- $Q(i) =$  conjunto dos índices que estão na próxima camada a camada do  $i$ -ésimo elemento e estão interligados ao  $i$ -ésimo elemento.

O algoritmo retropropagação é abordado na literatura sob várias formas com o propósito de torná-lo mais rápido computacionalmente. Uma formulação bastante interessante é o algoritmo retropropagação com momento [19].

Então, efetuando-se o cálculo do gradiente como indicado na equação (3), considerando-se as funções de ativação e o termo momento, obtém-se o seguinte esquema de adaptação de pesos [19]:

$$v_{ij}(h+1) = v_{ij}(h) + Dv_{ij}(h) \quad (7)$$

sendo:

- $Dv_{ij}(h) = 2g(1-h) b_i x_i + h Dv_{ij}(h-1)$ ;
- $v_{ij} =$  peso correspondente à interligação entre o  $i$ -ésimo e  $j$ -ésimo neurônio;
- $h =$  constante momento ( $0 \leq h < 1$ ).

O parâmetro  $b_j$  é calculado de forma semelhante ao esquema de adaptação realizado pelo algoritmo retropropagação convencional sem momento.

## Funções de Ativação

As funções de ativação de um neurônio artificial são funções diferenciáveis e não-lineares. Esta não linearidade se torna importante pelo fato desta função de ativação fornecer a rede neural uma saída similar a saída de um neurônio, que normalmente não é linearmente dependente de suas entradas.

Nesta seção serão apresentadas as funções de ativação utilizadas para realizar o treinamento da rede neural artificial, mostradas na Tabela 1. Dentre elas, as funções de base radial [4], [7], [10], [11], [14]: Gaussiana (G), Multiquadrática Recíproca (MR), Multiquadrática Recíproca Inversa (MRI) e Secante Hiperbólica (SH) e função Sigmóide (S) [8].

Tabela 1: Funções de ativação.

	G	MR	MRI	SH	S
<b>Função</b>	$y_i = e^{-cx^2}$	$y_i = \frac{1}{\sqrt{(x^2 + c^2)}}$	$y_i = \frac{1}{c} - \frac{1}{\sqrt{x^2 + c^2}}$	$y_i = \frac{2}{e^{cx} + e^{-cx}}$	$y_i = \frac{1}{1 + e^{-Ix}}$
<b>Derivada</b>	$y_i' = -2c y_i$	$y_i' = -\frac{x}{(x^2 + c^2)} y_i$	$y_i' = \frac{x}{(x^2 + c)^{3/2}}$	$y_i' = -\frac{2}{(e^{cx} + e^{-cx})^2 (ce^{cx} - ce^{-cx})}$	$y_i' = I y_i (1 - y_i)$

## Aplicação

A previsão de carga elétrica, a curto-prazo (previsão diária), é realizada da seguinte forma:

- A implementação de uma recorrência na saída de um determinado instante é usada como entrada para o instante subsequente.
- A saída da rede corresponde ao valor da carga relativo à hora ( $h+1$ ).

São considerados os dados históricos horários dentro de um intervalo preestabelecido, *e.g.*, mensal. A entrada da rede neural, para uma determinada hora  $h$ , é definida como sendo os valores da carga, extraídos dos dados históricos em quatro instantes (o valor corrente, uma, duas e três horas anteriores), e dados referentes ao tempo (dia da semana e hora).

O conjunto que compõe a entrada / saída é definido considerando esta estratégia até completar todo o intervalo de tempo da série temporal. Este esquema pode ser modificado visando obter melhores resultados, através da introdução de outras variáveis (temperatura, dias nublados, etc.). Assim, o vetor de entrada e a saída, respectivamente, são definidos da seguinte forma [9]:

$$\mathbf{X}(h) = [ \mathbf{t}^T \ L(h-3) \ L(h-2) \ L(h-1) \ L(h) ]^T, \quad \mathbf{X} \in R^m$$

$$\mathbf{Y}(h) = [ L(h+1) ], \quad \mathbf{Y} \in R^l$$

sendo:

$m$  = dimensão do vetor  $\mathbf{X}$ ;

$L(h-p)$  = valor da carga  $p$  horas anterior à hora corrente  $h$ ;

$L(h+1)$  = valor da carga elétrica correspondente à hora subsequente a hora corrente  $h$ ;

$\mathbf{t}$  = vetor de tempo referente aos dados históricos (mês, dia da semana, feriado, hora, etc) representados de modo similar ao código binário  $(-1,+1)$ .

A escolha desta representação binária é preferível em relação à representação  $(0,+1)$ , tendo em vista que

o componente “0” de entrada à rede não proporciona modificação nos pesos. Deste modo, a representação  $(-1,+1)$  produz uma convergência mais rápida, consequentemente sendo mais eficiente.

As cargas elétricas  $L(h-3), \dots, L(h-1)$  representam o laço de realimentação, com atraso, da saída. Deste modo, esta rede é do tipo recorrente.

Os dados históricos são compostos por valores de cargas elétricas obtidas de uma companhia do setor elétrico brasileiro. Estes dados contêm cargas horárias referentes ao ano de 1998, sendo relatados os dias não-típicos (feriados), dias especiais (sábados e domingos) e dias de uma semana típica. É importante enfatizar que o dados que compõem o banco de entrada para o treinamento não contêm dias atípicos.

O esquema de previsão de cargas elétricas utilizado é efetuado para dois sistemas previsores:

- O primeiro sistema (**Sistema Previsor 1 – SP1**) contêm dados históricos compreendidos entre 8 a 28 de julho de 1998 e dados relativos ao tempo (dia da semana e a hora) referentes a carga horária utilizada durante o período. Portanto, o banco de dados está representado por um vetor de 504 entradas com suas respectivas saídas que correspondem efetivamente a dados de cargas horárias fornecidas em um período de 21 dias.

- O segundo sistema (**Sistema Previsor 2 – SP2**) contêm dados compreendidos do dia 17 de junho a 28 de julho de 1998 e dados de tempo (dia da semana e a hora) relativos a carga horária utilizada durante o período. Logo, tem-se que este banco de dados é denotado por um vetor de 1008 entradas e respectivas saídas referentes a um período que abrange 42 dias.

- O terceiro sistema (**Sistema Previsor 3 – SP3**) contêm dados compreendidos do dia 4 de maio a 27 de julho de 1998 e dados de tempo (dia da semana e a hora) relativos a carga horária utilizada durante o período. Este banco de dados compõe um vetor de 2016 padrões de entradas com suas respectivas saídas abrangendo um período de 84 dias.

Tabela 2: Especificação dos parâmetros da rede neural.

Funções de Ativação		Parâmetros				
		NNC	Tol	$\gamma$	$\eta$	$\lambda$ ou $c$
SP1	G	12-50-1	0,035	1,10	0,70	0,50
	MR	12-15-1	0,030	0,99	0,90	0,50
	MRI	12-20-1	0,029	0,50	0,90	1,25
	SH	12-50-1	0,029	1,10	0,90	0,90
	S	12-30-1	0,020	4,50	0,90	0,30
SP2	G	12-15-1	0,029	1,10	0,70	0,40
	MR	12-45-1	0,029	0,90	0,75	1,30
	MRI	12-20-1	0,029	0,82	0,90	1,20
	SH	12-55-1	0,030	1,10	0,90	1,00
	S	12-30-1	0,025	5,50	0,90	0,30
SP3	G	12-20-1	0,029	0,50	0,90	0,65
	MR	12-25-1	0,029	0,90	0,73	0,65
	MRI	12-20-1	0,030	1,10	0,70	1,20
	SH	12-15-1	0,030	1,10	0,90	1,00
	S	12-33-1	0,030	6,50	0,70	0,50

A Tabela 2 apresenta os principais parâmetros utilizados no treinamento e diagnóstico da rede neural utilizando as cinco funções de ativação para os três sistemas previsores. Na Tabela 2 tem-se que: NNC é número de neurônios por camada, Tol é a tolerância,  $\lambda$  é a inclinação da função sigmóide e  $c$  é o centro das funções de base radial.

Para efetuar a análise da precisão dos resultados, foram utilizados: erro percentual médio absoluto (MAPE) [17] e o erro máximo (EM) da previsão diária, assim definidos:

$$MAPE = \frac{1}{N_t} \left\{ \sum_{i=1}^N (|L(h) - \underline{L}(h)| / L(h)) \right\} \times 100 \quad (15)$$

$$Erro \text{ Máximo } (\%) = \max \{ |L(h) - \underline{L}(h)| / L(h) \} \times 100 \quad (16)$$

em que:

$L(h)$  = valor real da carga referente à hora  $h$ ;  
 $\underline{L}(h)$  = valor estimado da carga referente à hora  $h$ ;  
 $N_t$  = número total de horas.

De forma a analisar o desempenho do processamento da rede neural aplicadas a diferentes funções de ativação, é definido o ganho de tempo de execução (processamento), considerando o tempo necessário para efetuar o treinamento e o diagnóstico, da seguinte forma:

$$GT = TRPFG / TRPFA$$

sendo:

$TRPFG$  = tempo de execução do algoritmo RP utilizando a função gaussiana (s);

$TRPFA$  = tempo de execução do algoritmo RP usando as demais funções de ativação (s).

O tempo de processamento diz respeito apenas à execução do algoritmo, sendo excluídas as operações de leitura e saída dos dados.

Foram realizados testes para os três sistemas previsores, utilizando as funções: Gaussiana, Multiquadrática Recíproca, Multiquadrática Recíproca Inversa, Secante Hiperbólica e Sigmóide. Dentre os resultados obtidos, para os três sistemas previsores, é apresentado, graficamente, apenas os melhores resultados.

O melhor resultado da previsão de carga para o SP1, referente ao dia 29 de julho de 1998, foi obtido pela função Secante Hiperbólica. A Figura 1 ilustra a curva real da carga e a curva obtida através da utilização da função Secante Hiperbólica.

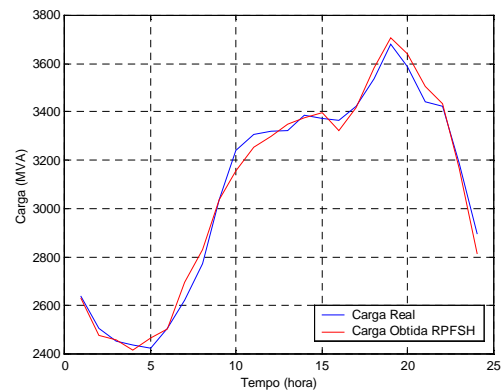


Figura 1: Resultado para previsão de carga para o sistema SP1 utilizando a função secante hiperbólica.

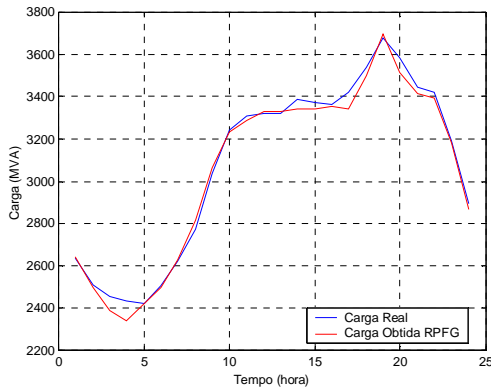


Figura 2: Resultado para previsão de carga para o sistema SP2 utilizando a função gaussiana.

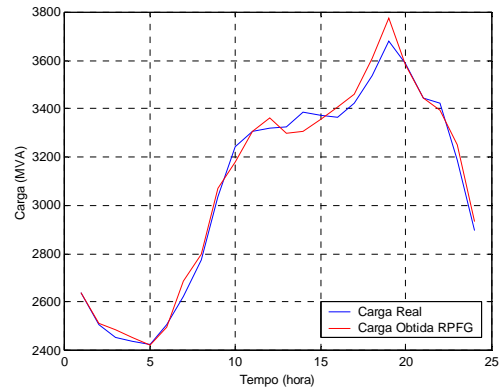


Figura 3: Resultado para previsão de carga para o sistema SP3 utilizando a função gaussiana.

Tabela 3: Resultados comparativos.

Funções de Ativação	Parâmetros					
	NI	TP (s)	GT	MAPE	EM	
SP1	G	524	11	1,00	1,30	4,46
	MR	1061	7	1,57	1,30	3,69
	MRI	2225	40	0,28	1,28	2,72
	SH	<b>1645</b>	<b>57</b>	<b>0,19</b>	<b>1,11</b>	<b>2,85</b>
	S	1878	33	0,33	1,35	3,98
SP2	G	<b>9442</b>	<b>110</b>	<b>1,00</b>	<b>0,96</b>	<b>3,89</b>
	MR	1162	43	2,56	1,06	3,10
	MRI	661	25	4,40	1,16	3,62
	SH	3228	243	0,45	1,10	4,00
	S	1019	39	2,81	1,12	3,45
SP3	G	<b>30333</b>	<b>1042</b>	<b>1,00</b>	<b>1,03</b>	<b>2,57</b>
	MR	68379	2048	0,51	1,26	3,63
	MRI	2193	163	6,39	1,16	3,81
	SH	20880	800	1,30	1,24	4,74
	S	5253	231	4,50	1,22	3,39

Já, para o SP2, o melhor resultado da previsão de carga, referente ao dia 29 de julho de 1998, foi obtido pela função Gaussiana ilustrada pela Figura 2. Do mesmo modo, para o SP3, o melhor resultado, referente ao dia 28 de julho de 1998, foi obtido utilizando a função Gaussiana que é apresentado através da Figura 3.

Os resultados comparativos de ambos os métodos estão relacionados na Tabela 3, no qual NI é o número de iterações e TP o tempo de processamento.

## Conclusão

Este artigo propôs investigar a utilização de novas funções de ativação (funções de base radial: gaussiana, multiquadrática recíproca, multiquadrática recíproca

inversa, secante hiperbólica) no treinamento e diagnóstico das redes neurais baseadas no algoritmo retropropagação.

O objetivo é aplicar o algoritmo retropropagação na resolução do problema de previsão de cargas elétricas a curto-prazo.

Através da Tabela 3, pode-se notar que para os três sistemas previsores, o algoritmo retropropagação, utilizando funções de base radial, obteve maior precisão, se comparado ao algoritmo retropropagação usando a função sigmóide. Pode-se observar também, que algumas funções de base radial, obtiveram um bom desempenho em termos de tempo de processamento para os sistemas previsores 1, 2 e 3.

Para os três sistemas previsores, pode-se notar, através de um estudo feito para as diversas funções de

base radial, que a utilização deste tipo de funções se sobressai ao treinamento executado com a função sigmóide. Pode-se facilmente observar que as funções de base radial, independente do sistema previsor, fornecem resultados mais precisos comparado a função sigmóide, apesar de em alguns casos ter tempo de processamento mais elevado.

## Referências

- [1] E. A. Bender, "Mathematical methods in artificial intelligence", IEEE Computer Society Press, Los Alamitos, California, 1996.
- [2] G. E. Box and G. M. Jenkins, "Times series analysis: forecasting and control", Holden-Day, San Francisco, 1976.
- [3] T. M. O'Donovan, "Short term forecasting: an introduction to the Box-Jenkins approach", John Wiley & Sons, New York, USA, 1983.
- [4] C. M. T. T. Fernandes and V. M. A. Leitão, Utilização de funções de base radial em problemas unidimensionais de análise estrutural, *Métodos Numéricos en Ingeniería V*, Madrid, Espanha, 2002, pp. 338-354.
- [5] T. L. Fine, "Feedforward Neural Network Methodology", Springer-Verlag, USA, 1999.
- [6] S. Haykin, "Redes neurais, princípios e prática", Bookman, Porto Alegre, 2001.
- [7] F. Heimes and B. V. Heuveln, The Normalized Radial Basis Function Neural Network, *IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 2, 1998, pp. 1609-1614.
- [8] S. V. Kartalopoulos, "Understanding neural networks and fuzzy logic", IEEE Press, New York, 1996.
- [9] M. L. M. Lopes, C. R. Minussi and A. P. Lotufo, Electrical load forecasting formulation by a fast neural network, *Engineering Intelligent Systems*, Vol. 11, No. 1, 2003, pp. 51-57.
- [10] N. Mai-Duy and T. Tran-Cong, Approximation of Function and its Derivatives Using Radial Basis Function Networks, *Elsevier Applied Mathematical Modelling*, Vol. 27, 2003, pp. 197-220.
- [11] J. F. Marar, Polinômios potências de sigmóide (PPS): Uma nova técnica para aproximação de funções, construção de Wavenets e suas aplicações em processamento de imagens e sinais, Tese de Doutorado, 1997.
- [12] C. R. Minussi and W. Freitas Filho, Sensitivity analysis for transient stability, *IEEE Proceedings on Generation, Transmission And Distribution*, Vol. 145, No. 6, 1998, pp. 669-674.
- [13] C. R. Minussi and M. C. G. Silveira, Transient stability analysis of electric power systems by neural networks, Rio de Janeiro, RJ, *38<sup>th</sup> Midwest Symposium on Circuits And Systems*, 1995, pp. 1305-1308.
- [14] B. Mulgrew, Applying Radial Basis Function, *IEEE Signal Processing Magazine*, Vol. 13, No. 2, 1996, pp. 50-65.
- [15] F. Shimit, M. I. M. S. Bueno and R. J. Poppi, Aplicação de alguns modelos quimiométricos à espectroscopia de fluorescência de raios-x de energia dispersiva, *Química Nova On Line*, Vol. 25, No. 6, 2002, pp. 949-956.
- [16] P. K. Simpson, "Artificial neural systems: foundations, paradigms, applications and implementations", Pergamon Press, New York, 1989.
- [17] D. Srinivasan, S. S. Tan, C. S. Chang and E. K. Chan, Practical implementation of a hybrid fuzzy neural network for one-day-ahead load forecasting, *IEE Proceedings Generation, Transmission and Distribution*, Vol. 145, No. 6, 1998, pp. 687-692.
- [18] P. J. Werbos, "Beyond regression: new tools for prediction and analysis in the behavioral sciences", Master Thesis, Harvard University, 1974.
- [19] B. Widrow and M. A. Lehr, 30 years of adaptive neural networks: perceptron, madaline, and backpropagation, *Proceeding of the IEEE*, Vol. 78, No. 9, 1990, pp. 1415-1442.