Técnicas de Regularização de Modelos Neurais Aplicadas à Previsão de Carga a Curto Prazo

Vitor Hugo Ferreira
Laboratório de Sistemas de Potência, Programa de Engenharia Elétrica, COPPE/UFRJ, 21945-970, Rio de Janeiro, RJ, Brasil.
vitor@vishnu.coep.ufrj.br

Alexandre P. Alves da Silva
Laboratório de Sistemas de Potência, Programa de Engenharia Elétrica, COPPE/UFRJ, 21945-970, Rio de Janeiro, RJ, Brasil.
alex@coep.ufrj.br

Abstract

The knowledge of loads' future behavior is very important for decision making in power system operation. During the last years, many load models have been proposed, and the neural ones have presented the best results. One of the disadvantages of the neural models is the possibility of excessive adjustment of the training data, named overfitting, degrading the generalization performance of the estimated models. This problem can be tackled by using regularization techniques. The present shows the application of some of these techniques to short term load forecasting.

1 Introdução

Muitas das decisões a serem tomadas na operação de sistemas de potência, tais como comissionamento de unidades geradoras, despacho econômico, controle automático da geração e elaboração de planos de manutenção, dependem do conhecimento prévio do comportamento da carga [1]. Diante disso, vários modelos vêm sendo propostos ao longo dos anos para prever a carga, com as estruturas baseadas em sistemas inteligentes, incluindo as chamadas redes neurais artificiais (RNA's) e sistemas baseados em lógica fuzzy; em conjunto com os modelos híbridos, apresentando os melhores resultados para o horizonte de curto prazo [3].

Entre os modelos baseados em redes neurais, os perceptrons de múltiplas camadas (MLPs) têm se mostrado bastante eficientes na tarefa de modelar a carga para o horizonte de curto prazo, que varia de poucos minutos até 168 horas à frente. Uma das vantagens destes modelos reside no teorema da aproximação universal, que demonstra que estes modelos, com uma única camada oculta contendo um número suficiente de neurônios, podem aproximar com precisão arbitrária qualquer função contínua não-linear. Entretanto, na presença de dados ruidosos, esta vantajosa característica das RNA's pode se tornar prejudicial, visto que estas, ao invés de representar o comportamento regular da carga, podem modelar o ruído, o chamado overfitting, resultando na degradação do desempenho para novos dados que não aqueles utilizados no treinamento. Modelos com esta característica apresentam pequena capacidade de generalização, devido principalmente à complexidade excessiva, isto é, número elevado de parâmetros livres.

O exposto acima ilustra a necessidade de controle da complexidade das RNA's, com o intuito de obter a melhor capacidade de generalização possível. Neste trabalho são aplicadas algumas técnicas de regularização na estimação de RNA's para previsão de carga a curto prazo, mais especificamente o treinamento Bayesiano, o escalonamento do ganho da função de ativação e as máquinas de vetor suporte (SVM).

Para ilustração das metodologias em estudo, são desenvolvidos modelos para previsão da carga horária e do pico de carga diário, utilizando duas bases de dados distintas. Para verificação da qualidade das estruturas obtidas, são realizadas previsões da curva de carga diária, ou seja, 24 horas à frente para o ano de 1991, para a série de carga em base horária, e previsões da curva mensal de pico de carga diário, isto é, 31 dias à frente, para janeiro de 1999, para a série de pico de carga diário.

2 Redes Neurais Artificiais

As RNA's comumente utilizadas em previsão de carga a curto prazo são do tipo *feedforward* com uma única camada oculta de neurônios. As subseções a seguir descrevem os modelos neurais do tipo MLP e SVM.

2.1 Perceptron de Múltiplas Camadas (MLP)

Seja $\underline{x} \in \mathbb{R}^n$ o vetor contendo os sinais de entrada, $d \in \mathbb{R}$ a saída desejada, $\underline{w} \in \mathbb{R}^M$, onde $M = mn + 2m + 1$ (M igual ao número total de parâmetros do modelo e m o número de neurônios na camada oculta) o vetor contendo os pesos das conexões e os bias b_k e b, $k = 1, 2, ..., m$, $\varphi(x) : \mathbb{R} \rightarrow \mathbb{R}$, uma função não-linear (para este trabalho $\varphi(x) = \tanh(x)$), e $y \in \mathbb{R}$, $y = f(\underline{x}, \underline{w})$, a saída gerada pela rede. Para o caso do MLP, a saída da rede é dada por:

$$c_k = \varphi\left[\sum_{j=1}^{n}\left(w_{kj} x_j\right) + b_k\right], k = 1, 2, ..., m \tag{1}$$

$$y = f(\underline{x}, \underline{w}) = \sum_{i=1}^{m} w_i c_i + b$$

Dado um conjunto D de N pares entrada-saída, $D = \{\underline{x}_i, d_i\}$, $i = 1, 2, ..., N$, o objetivo do treinamento de um MLP reside na estimativa do vetor de pesos $\underline{w}$ que minimize o risco empírico dado por:

$$E_s(\underline{w}, D) = \frac{1}{2}\sum_{j=1}^{N}\left[d_i - f(\underline{x}_i, \underline{w})\right]^2 = \frac{1}{2}\sum_{j=1}^{N}\left[d_i - y_i\right]^2 \tag{2}$$

A principal desvantagem dos métodos de treinamento que minimizam o funcional descrito em (2) é a possibilidade de ocorrência de *overfitting.*

2.2 Máquina de Vetor Suporte (SVM)

Para as máquinas de vetor suporte, a saída da rede é dada por:

$$y = \sum_{j=0}^{m} W_j \phi_j(\underline{x}) = \underline{W}^T \underline{\phi}(\underline{x}) \tag{3}$$

$$\underline{\phi}(\underline{x}) = \left[1, \phi_1(\underline{x}), \phi_2(\underline{x}), ..., \phi_m(\underline{x})\right]^T$$

$$\underline{W} = \left[b, W_1, W_2, ..., W_m\right]^T$$

onde $\underline{\phi}(\underline{x})$ representa um conjunto de funções de base não-lineares definidas pelo usuário. Seja a função de perda, com tolerância $\varepsilon, L_\varepsilon(d, y)$, dada por [7]:

$$L_\varepsilon(d, y) = \begin{cases} |d - y| - \varepsilon, & |d - y| \geq \varepsilon \\ 0, & |d - y| < \varepsilon \end{cases} \tag{4}$$

Na equação (4), ε é um parâmetro especificado pelo usuário. O objetivo do treinamento de uma SVM é a minimização do risco empírico dado por:

$$E_s(\underline{W}, D) = \frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(d_i, y_i) \tag{5}$$

sujeito à restrição:

$$\|\underline{W}\|^2 \leq c_0 \tag{6}$$

onde c_0 responde pelo controle de complexidade do modelo [8].

A restrição não-linear representada na equação (6) pode ser abordada na função objetivo de um problema de programação quadrática dado por:

$$\min \Phi(\underline{W}, \underline{\xi}, \underline{\xi}') = C\left[\sum_{i=1}^{N}(\xi_i + \xi_i')\right] + \frac{1}{2}\underline{W}^T \underline{W} \tag{7}$$

onde $\underline{\xi} = \left[\xi_1, \xi_2, ..., \xi_N\right]^T$, $\underline{\xi}' = \left[\xi_1', \xi_2', ..., \xi_N'\right]^T$, sujeito às restrições:

$$\begin{aligned} & d_i - \underline{W}^T \underline{\phi}(\underline{x}_i) \leq \varepsilon + \xi_i \\ & \underline{W}^T \underline{\phi}(\underline{x}_i) - d_i \leq \varepsilon + \xi_i' \\ & \xi_i \geq 0, \quad \xi_i' \geq 0, \quad i = 1, 2, ..., N \end{aligned} \tag{8}$$

Na equação (7), C é um parâmetro especificado pelo usuário, responsável pelo equilíbrio entre o ajuste dos dados de treinamento e a complexidade do modelo.

Para solução deste problema de otimização, pode ser definida a função Lagrangeana, dada por:

$$\begin{aligned} J(\underline{W}, \underline{\xi}, \underline{\xi}', \underline{\alpha}, \underline{\alpha}', \underline{\gamma}, \underline{\gamma}') = {} & C\sum_{i=1}^{N}(\xi_i + \xi_i') + \frac{1}{2}\underline{W}^T \underline{W} \\ & - \sum_{i=1}^{N}\alpha_i\left[\underline{W}^T \underline{\phi}(\underline{x}_i) - d_i + \varepsilon + \xi_i\right] \\ & - \sum_{i=1}^{N}\alpha_i'\left[d_i - \underline{W}^T \underline{\phi}(\underline{x}_i) + \varepsilon + \xi_i'\right] \\ & - \sum_{i=1}^{N}(\gamma_i \xi_i + \gamma_i' \xi_i') \end{aligned} \tag{9}$$

$$\underline{\alpha} = \left[\alpha_1, \alpha_2, ..., \alpha_N\right]^T, \quad \underline{\alpha}' = \left[\alpha_1', \alpha_2', ..., \alpha_N'\right]^T$$

$$\underline{\gamma} = \left[\gamma_1, \gamma_2, ..., \gamma_N\right]^T, \quad \underline{\gamma}' = \left[\gamma_1', \gamma_2', ..., \gamma_N'\right]^T$$

onde $\underline{\alpha}$ e $\underline{\alpha}'$ são os multiplicadores de Lagrange. Das condições de otimalidade do cálculo, são encontradas as seguintes equações:

$$\begin{aligned} & \underline{W} = \sum_{i=1}^{N}(\alpha_i - \alpha_i')\underline{\phi}(\underline{x}_i) \\ & \gamma_i = C - \alpha_i, \quad \gamma_i' = C - \alpha_i', \quad i = 1, 2, ..., N \end{aligned} \tag{10}$$

De posse das equações (10), o problema dual de maximização, correspondente ao problema primal de minimização da função definida em (9), pode ser formulado como:

$$\max Q(\underline{\alpha},\underline{\alpha}') = \sum_{i=1}^{N} d_i(\alpha_i - \alpha_i') - \varepsilon \sum_{i=1}^{N}(\alpha_i + \alpha_i') \tag{11}$$

$$-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i')(\alpha_j - \alpha_j')K(\underline{x}_i, \underline{x}_j)$$

sujeito às restrições:

$$\sum_{i=1}^{N}(\alpha_i - \alpha_i') = 0 \tag{12}$$

$$0 \le \alpha_i \le C, \quad 0 \le \alpha_i' \le C, \quad i = 1, 2, ..., N$$

Na equação (11), $K(\underline{x}_i, \underline{x}_j) = \underline{\phi}^T(\underline{x}_i)\underline{\phi}(\underline{x}_j)$ é o núcleo do produto interno (kernel), definido conforme o teorema de Mercer [7]-[9]. Neste trabalho, será utilizado o *kernel* gaussiano. Portanto, a saída da máquina de vetor suporte é dada por:

$$y = f(\underline{x}, \underline{W}) = \sum_{i=1}^{N}(\alpha_i - \alpha_i')K(\underline{x}, \underline{x}_i) \tag{13}$$

Como pode ser visto através da equação (13), os vetores de suporte são definidos como os padrões do conjunto de treinamento que apresentem $\alpha_i \neq \alpha_i'$, ou seja, aqueles situados fora da banda definida por ε. Isto é equivalente a dizer que uma SVM é um modelo feedforward com função de ativação definida pelo kernel $K(\underline{x}, \underline{x}_i)$.

2.3 MLP versus SVM

Do exposto acima, algumas diferenças importantes podem ser observadas entre o MLP e a SVM. No treinamento do MLP via retro-propagação de erro, uma estrutura para o mesmo (número de neurônios na camada intermediária) deve ser definida a priori. Nas SVMs, a estrutura do modelo é resultado do algoritmo de treinamento, dependente dos parâmetros ε, C e do tipo de *kernel* selecionado. O treinamento do MLP via retro-propagação de erro é baseado na minimização do risco empírico. Já o das SVM é baseado na minimização do risco estrutural, princípio que busca a minimização do limite superior do erro de generalização. Além disso, enquanto que para o MLP a superfície a ser otimizada é extremamente não-convexa, repleta de mínimos locais, para a SVM esta superfície apresenta forma quadrática com um único ponto de máximo. Consequentemente, para um mesmo conjunto de dados e valores de ε e C constantes, a solução para a SVM é única.

3 Técnicas de Regularização

Conforme mencionado na seção 2.1, o treinamento de um MLP através da minimização do risco empírico pode levar à ocorrência de overfitting, evidenciando a necessidade do controle de complexidade do MLP.

Existem dois procedimentos gerais para o controle da complexidade de MLPs. O primeiro é a chamada estabilização de estrutura [10], onde o objetivo reside na determinação do número suficiente de neurônios na camada oculta, que pode ser implementado de três formas.

Uma delas consiste em comparar modelos com números de neurônios diferentes na camada intermediária, escolhendo a estrutura que apresentar melhor desempenho em um conjunto independente de dados.

De outra forma, o processo de modelagem pode ser iniciado através de um modelo demasiadamente complexo, o qual é submetido a alguns algoritmos de poda. Por último, tal processo pode ser iniciado com uma rede excessivamente simples, sendo adicionados neurônios ao longo do treinamento. Aplicações destas variantes podem ser encontradas em [13].

Outro procedimento utilizado para controlar a complexidade das RNA's tem por base a teoria da regularização. Neste procedimento, o compromisso entre o ajuste dos dados de treinamento e a capacidade de generalização pode ser obtido através da minimização do risco total, expresso por:

$$R(\underline{w}) = E_s(\underline{w}, D) + \lambda E_c(\underline{w}) \tag{14}$$

Na equação acima, o primeiro termo representa o ajuste do modelo aos dados de treinamento (risco empírico). O segundo termo está relacionado com a complexidade do modelo, impondo à solução conhecimento prévio acerca desta, podendo adquirir diversas formas, conforme apresentado em [10], [11] e [13]. O parâmetro λ é conhecido como parâmetro de regularização, o qual define a importância relativa entre o ajuste dos dados de treinamento e a complexidade do modelo. Uma das formas para o funcional $E_c(\underline{w})$ é obtida através da aplicação de técnicas de inferência Bayesiana na determinação do vetor de parâmetros $\underline{w}$ que define o MLP [11]. Pela regra de Bayes, a função densidade de probabilidade do vetor $\underline{w}$, $p(\underline{w} | D)$, dado o conjunto de pares entrada-saída D, é dada por:

$$p(\underline{w} | D) = \frac{p(D | \underline{w})\, p(\underline{w})}{p(D)} \tag{15}$$

Na equação (15), $p(D | \underline{w})$ representa a função densidade de probabilidade de ocorrência do conjunto D, dado o vetor de parâmetros $\underline{w}$, $p(\underline{w})$ representa a função densidade de probabilidade a priori de $\underline{w}$, e $p(D) = \int p(D | \underline{w})\, p(\underline{w})\, d\underline{w}$ representa um fator de normalização, que garante $\int p(\underline{w} | D)\, d\underline{w} = 1$.

Assumindo que o vetor $\underline{w}$ apresenta a priori uma distribuição Gaussiana com vetor média nulo e matriz de covariância $\alpha^{-1}\underline{\underline{I}}$, onde $\underline{\underline{I}}$ é a matriz identidade de dimensão MxM, e que as saídas desejadas d_i são dadas por $d_i = f(\underline{x}_i, \underline{w}) + \zeta$, onde ζ é um ruído gaussiano de média nula e variância β^{-1}, maximizar a distribuição de probabilidade a posteriori de $\underline{w}$ é equivalente a minimizar a seguinte expressão [11]:

$$S(\underline{w}) = \frac{\beta}{2}\sum_{i=1}^{N}\left[f(x_i, \underline{w}) - d_i\right]^2 + \frac{\alpha}{2}\sum_{i=1}^{M} w_i^2 \tag{16}$$

Como o único ponto de interesse em $S(\underline{w})$ é o ponto de mínimo, fazendo $\lambda = \alpha/\beta$, $S(\underline{w}) = R(\underline{w})$, e, portanto,

$$E_c(\underline{w}) = \frac{1}{2}\|\underline{w}\|^2 \tag{17}$$

A vantagem da abordagem Bayesiana está baseada no fato que esta fornece um mecanismo iterativo para estimativa ao longo do treinamento tanto de α quanto de β e, portanto, de λ [11]. Este tipo de treinamento de MLPs é conhecido como treinamento Bayesiano.

Redes com boa capacidade de generalização podem ser obtidas através de outros métodos, onde a complexidade do modelo não é controlada diretamente, ou seja, $\lambda = 0$. Dentre estes métodos, podem ser citados a parada antecipada do treinamento, o treinamento com inserção de ruído e o escalonamento do ganho da função de ativação [14].

Na parada antecipada do treinamento, validação cruzada é utilizada para particionar o conjunto de treinamento em dois subconjuntos, um de estimação de $\underline{w}$ e outro de validação do modelo. A atualização do vetor de pesos é interrompida quando o erro para o conjunto de validação sofrer deterioração em relação às iterações anteriores. Apesar da elevada aplicabilidade deste procedimento, vide [5], [13], [15] e [16], a teoria estatística recomenda cautela na aplicação desta técnica, conforme apresentado em [12].

O treinamento com inserção de ruído consiste em adicionar ao conjunto de treinamento versões corrompidas dos padrões originalmente pertencentes a este. Esses padrões são gerados através da adição de ruído nas variáveis de entrada. Desta forma, para padrões de entrada similares, a saída sofrerá pouca ou nenhuma alteração, o que é equivalente a supor que a função a ser aproximada apresenta um certo grau de suavidade, conforme assumido na teoria de regularização de *Tikhonov* [14].

Segundo [14], para RNA's com uma única camada oculta contendo neurônios não-lineares e uma única saída linear, se as amostras do conjunto de treinamento forem obtidas seguindo uma distribuição uniforme e o ruído adicionado às entradas for Gaussiano com vetor média nulo e matriz de covariância $\sigma^2_{ruido}\underline{\underline{I}}$ (neste caso $\underline{\underline{I}}$ tem dimensão nxn), a rede obtida através da minimização do risco empírico irá apresentar capacidade de generalização similar à rede treinada com o conjunto original de dados (não corrompido) utilizando uma outra técnica de regularização. Da mesma forma, uma rede treinada com o conjunto original de dados, porém sem utilizar um regularizador, ou seja, $\lambda = 0$, irá apresentar capacidade de generalização similar às regularizadas se os ganhos das funções de ativação dos neurônios da camada oculta forem multiplicados pelo fator a_k dado por:

$$a_k = \frac{1}{\sqrt{\|\underline{w}_k\|^2 \sigma^2_{ruido} + 1}} \tag{18}$$

$$\underline{w}_k = \left[w_{k1}, w_{k2}, ..., w_{kn}\right]^T,\ k = 1, 2, ..., m$$

4 Pré-processamento dos Dados

Como é de conhecimento amplo, o pré-processamento dos dados constitui uma das principais ferramentas de melhoria de desempenho de qualquer modelo estimado por técnicas de identificação de sistemas. Logo, algumas transformações foram efetuadas às bases de dados originais, as quais serão apresentadas a seguir.

4.1 Previsão da Carga Horária

Para desenvolvimento dos modelos de previsão da carga horária, foram utilizadas as séries de carga e temperatura da Puget Sound Power and Light Company, as quais podem ser obtidas no endereço eletrônico http://www.ee.washington.edu /class/559/2002spr.

O primeiro tratamento aplicado às séries está relacionado com os dados faltantes. Para corrigir este problema, estes valores foram substituídos pelas respectivas médias aritméticas entre o valor anterior ao dado faltante e o valor posterior a este.

Em seguida foram retiradas da série de carga as sazonalidades diária e semanal, além da aplicação da primeira diferença, com o intuito de remover a tendência da série em questão. A seqüência de transformações foi a seguinte:

$$\begin{aligned} Serie1(k) &= Serie(k) - Serie(k-1) \\ Serie2(k) &= Serie1(k) - Serie1(k-24) \\ Serie3(k) &= Serie2(k) - Serie2(k-168) \end{aligned} \tag{19}$$

A série *Serie3* foi reduzida (média nula e variância unitária) e posteriormente normalizada no intervalo [-1,1], gerando a série *S(k)*. Após a normalização, foi calculada a autocorrelação parcial desta série, com o intuito de determinar as entradas utilizadas pelos modelos, sendo verificado que os atrasos com correlação mais significativa são *S(k-1), S(k-2), S(k-24) e S(k-168)*.

Após a retirada da sazonalidade diária da série de temperatura e posterior normalização da mesma, foi verificado, através da correlação cruzada entre esta série e *S(k)*, que os atrasos mais significativos são *T(k), T(k-1), T(k-2), T(k-24) e T(k-168)*.

Vale mencionar que, por comodidade, para realização das previsões de carga para as 24 horas de um determinado dia, as entradas *T(k), T(k-1) e T(k-2)* são alimentadas sempre com os respectivos valores medidos da série de temperatura (desazonalizada e normalizada), não sendo realizadas previsões de temperatura.

Além destes atrasos, são utilizadas também duas entradas contendo informações sobre a hora do dia, codificadas da seguinte forma [15], [16]:

$$HS(k) = sen\left(\frac{2\pi k}{24}\right); HC(k) = \cos\left(\frac{2\pi k}{24}\right) \tag{20}$$

$k = 1, 2, ..., 24$

Foram desenvolvidos diferentes modelos para cada dia da semana, incluindo finais de semana. O conjunto de treinamento de cada modelo consiste nos padrões referentes às últimas seis semanas do respectivo dia da semana. A Figura 1 apresenta um diagrama esquemático exemplificando o procedimento de construção do conjunto de treinamento e posterior previsão.

CONJUNTO DE TREINAMENTO						PREVISÃO
17/5/1990	24/5/1990	31/5/1990	7/6/1990	14/6/1990	21/6/1990	28/6/1990
24 HORAS	24 HORAS	24 HORAS	24 HORAS	24 HORAS	24 HORAS	24 HORAS

CONJUNTO DE TREINAMENTO						PREVISÃO
24/5/1990	31/5/1990	7/6/1990	14/6/1990	21/6/1990	28/6/1990	5/7/1990
24 HORAS	24 HORAS	24 HORAS	24 HORAS	24 HORAS	24 HORAS	24 HORAS

Figura 1: Arranjo do conjunto de treinamento e da previsão para um dia específico da semana

4.2 Previsão do Pico de Carga Diário

Os modelos para previsão do pico de carga diário foram desenvolvidos utilizando os dados de carga de uma empresa de energia européia, disponibilizados pelo EUNITE para realização de uma competição no ano de 2001. Este conjunto de dados pode ser encontrado no endereço http://neuron.tuke.sk/competition. Na competição, a tarefa dos competidores residia no desenvolvimento de modelos para previsão do pico de carga diário para todo o mês de janeiro de 1999. Desta forma, esta metodologia também será utilizada para avaliação dos modelos desenvolvidos neste trabalho.

Visto que não são disponibilizados dados de temperatura para janeiro de 1999, esta informação não será utilizada como entrada dos modelos. Além disso, já que devem ser realizadas previsões em base diária para todo o mês de janeiro, será utilizada a série de pico de carga diário obtida da série disponível discretizada em base de trinta minutos. Por último, ao contrário dos modelos desenvolvidos para previsão horária, a tendência e as sazonalidades existentes na série de carga não serão previamente tratadas.

Após a redução e normalização no intervalo [-1;1] da série de pico de carga diário, dando origem à série $P(k)$, foi estimada a função de autocorrelação parcial da mesma, a qual permitiu definir como entradas dos modelos os atrasos $P(k-1)$, $P(k-2)$,..., $P(k-7)$.

Para codificação do dia da semana, são utilizadas sete entradas binárias, sendo também utilizada uma oitava entrada binária para sinalização de feriados.

Diferentemente dos modelos de previsão de carga horária, será utilizada uma única estrutura para cada metodologia, já que a dependência do pico de carga com o dia da semana está representada no conjunto de entradas binárias. Em virtude da conhecida influência das estações do ano, serão utilizados para treinamento apenas dados referentes à estação do ano onde serão realizadas as previsões, ou seja, apenas os padrões referentes aos meses de janeiro a março e de outubro a dezembro.

5 Resultados

Para os modelos de previsão de carga horária, foram obtidos os resultados apresentados nas tabelas I e II. Na tabela I, são apresentadas as estruturas que apresentaram os melhores resultados, em termos do erro absoluto médio percentual de previsão para a curva de carga diária, para janeiro de 1991.

A tabela II apresenta os erros absolutos médios percentuais obtidos para o período de 1º. de fevereiro de 1991 a 31 de dezembro de 1991, sendo destacados os menores erros obtidos. A última coluna apresenta a diferença percentual entre o melhor e o pior resultado obtido para o respectivo dia.

Tabela I – Estrutura e parâmetros dos modelos de previsão de carga horária

	Sem Regularizador	*Bayesiano*	Escalonamento		SVM		
	Neurônios na camada oculta	Neurônios na camada oculta	Neurônios na camada oculta	σ^2_{ruido}	C	ε	σ_{kernel}
Segunda	2	2	2	0.12	0.1	0.100	4.24
Terça	2	2	2	0.16	0.1	0.001	4.24
Quarta	3	2	3	0.07	1.0	0.100	2.72
Quinta	3	2	2	0.17	1.0	0.400	1.96
Sexta	2	2	2	0.12	1.0	0.400	3.48
Sábado	4	2	4	0.05	0.1	0.001	5
Domingo	2	2	2	0.11	1.0	0.100	1.96

Os resultados obtidos para os modelos desenvolvidos para previsão do pico de carga diário estão apresentados na tabela III. A primeira linha desta tabela contêm os parâmetros dos modelos que apresentaram melhor desempenho para o período de dezembro de 1998. Utilizando estas estruturas, e incorporando ao conjunto de treinamento os dados referentes a dezembro de 1998, foram realizadas previsões para janeiro de 1999, dando origem aos resultados apresentados nas duas últimas linhas da tabela III.

Tabela II – Comparação entre os modelos de previsão da carga horária

	ARX	Sem Regularizador	*Bayesiano*	Escalonamento	SVM	Ganho de Desempenho
Segunda	8.23	8.76	6.43	7.00	5.23	40.3
Terça	8.16	7.04	7.47	6.52	4.97	39.1
Quarta	8.15	6.94	7.08	6.18	5.00	38.7
Quinta	8.15	10.21	6.93	8.41	7.83	32.1
Sexta	9.54	7.33	6.29	7.18	6.39	34.1
Sábado	7.38	9.57	7.38	8.17	5.24	45.3
Domingo	7.02	8.19	6.91	7.42	5.00	38.9
Média	8.09	8.29	6.93	7.27	5.66	31.7

Tabela III – Estruturas e resultados dos modelos de previsão de carga diária

	Sem Regularizador	*Bayesiano*	Escalonamento		SVM			Ganho de Desempenho
	Neurônios na camada oculta	Neurônios na camada oculta	Neurônios na camada oculta	σ^2_{ruido}	C	ε	σ_{kernel}	
	2	2	2	0.148	1	0.1	4.8	
Erro Médio[%]	2.47	2.02	7.53		2.36			73.2
Erro Máximo [%]	7.64	6.41	13.64		7.79			53.0

6 Conclusões

Neste trabalho foram aplicadas algumas técnicas de regularização de modelos neurais para previsão de carga a curto prazo. Os resultados apresentados evidenciam a potencialidade de tais técnicas, visto que estas, com exceção do escalonamento do ganho da função de ativação para previsão do pico de carga, deram origem a modelos com capacidade de generalização superior aos estimados através da aplicação do algoritmo de retropropagação do erro tradicional.

Neste contexto, deve ser salientado o desempenho da SVM, que apresentou os melhores resultados para cinco dos sete modelos desenvolvidos para previsão da carga horária. Mesmo com este desempenho superior, é esperado que os resultados obtidos possam ser melhorados através da busca de valores ótimos para os parâmetros que definem a SVM.

Também merece destaque o desempenho do treinamento bayesiano de MLPs, que deu origem ao modelo com melhor desempenho para previsão do pico de carga diário. Comparando com os modelos desenvolvidos na competição de 2001, o treinamento bayesiano apresentou desempenho equivalente ao obtido pelo vencedor desta, conforme pode ser verificado no endereço da competição (http://neuron.tuke.sk/competition.).

7 Agradecimentos

Os autores agradecem o apoio da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e do CNPq (Conselho Nacional de Desenvolvimento Científico e Pesquisa) pelo suporte financeiro.

8 Referências Bibliográficas

[1] A.S. Debs, *Modern Power Systems Control and Operation*, Kluwer Academic Publishers, 1988.

[2] D.W. Bunn, “Forecasting Loads and Prices in Competitive Power Markets”, *Proceedings of the IEEE*, v.88, n.2, pp. 163-169, Feb.2000.

[3] H.S. Hippert, C.E. Pedreira, “Neural Networks for Short-Term Load Forecasting: A Review and Evaluation”, *IEEE Transactions on Power Systems*, v.16, n.1, pp. 44-55, Feb. 2001.

[4] D.C. Park, M.A. El-Sharkawi, R.J. Marks II, “An Adaptively Trained Neural Network”, *IEEE Transactions on Neural Networks*, v.2, n.3, pp. 334-345, May 1991.

[5] A. Khotanzad, R. Afkhami-Rohani, D. Maratukulam, “ANNSTLF – Artificial Neural Network Short-Term Load Forecaster – Generation Three”, *IEEE Transactions on Power Systems*, v.13, n.4, pp. 1413-1422, Nov. 1998.

[6] S. Osowski, K. Siwek, “Regularization of Neural Networks for Improved Load Forecasting in the Power System”, *IEE Proceedings on Generation, Transmission and Distribution*, v.149, n.3, pp. 340-344, May 2002.

[7] V.N. Vapnik; *Statistical Learning Theory*, New York, John Wiley & Sons, 1998.

[8] V. Cherkassky, F. Mulier, *Learning from Data - Concepts, Theory and Methods*, John Wiley & Sons, New York, USA, 1998.

[9] B.Schölkopf, A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*, Cambridge, Massachusetts, 2002.

[10] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, 1995.

[11] D.J.C. Mackay, *Bayesian Methods for Adaptive Models*, Ph.D. dissertation, California Institute of Technology, Pasadena, California, USA, 1992.

[12] S. Amari, N. Murata, K.R. Müller, M. Finke, H. Yang, “Statistical Theory of Overtraining – Is Cross-validation Asymptotically Effective?”, *Advances in Neural Information Processing Systems*, v.8, pp. 176-182, 1996.

[13] N.K. Treadgold, T.D. Gedeon, “Exploring Constructive Cascade Networks”, *IEEE Transactions on Neural Networks*, v.10, n.6, 1335-1350, 1999.

[14] R. Reed., R.J. Marks II, S. Oh, “Similarities of Error Regularization, Sigmoid Gain Scaling, Target Smoothing and Training with Jitter”, *IEEE Transactions on Neural Networks*, v.6, n.3, 529-538, 1995.

[15] A.P. Alves da Silva, L.S. Moulin, “Confidence Intervals for Neural Network Based Short-Term Load Forecasting”, *IEEE Transactions on Power Systems*, v.15, n.4, 1191-1196, 2000.

[16] A.J.R. Reis, A.P. Alves da Silva, “Feature Extraction Via Multi-Resolution Analysis for Short-Term Load Forecasting”, *IEEE Transactions on Power Systems*, v.20, issue 1, pp. 189-198, Feb. 2005