

SELEÇÃO DE VARIÁVEIS *STEPWISE* APLICADAS EM REDES NEURAIS ARTIFICIAIS PARA PREVISÃO DE DEMANDA DE CARGAS ELÉTRICAS

MARLEIDE F. ALVES¹, ANNA DIVA P. LOTUFO¹, MARA LÚCIA M. LOPES²

1. *Laboratório de Sistemas Inteligentes, Departamento de Engenharia Elétrica, UNESP*
Av. Brasil 56, PO Box 31, CEP: 15385-000 – Ilha Solteira – SP
marleide_alves@bol.com.br, annadiva@ieee.org

2. *Departamento de Matemática, UNESP*
Rua Rio de Janeiro 266, CEP: 15385-000 – Ilha Solteira – SP
mara@mat.feis.unesp.br

Abstract – Load forecasting is very important for generation and electrical distribution planning. This work presents a hybrid model for short term load forecasting using stepwise selection algorithm aided by multiple linear regression and multilayer Perceptron neural network by backpropagation training to adjust the synaptic weights. Results are very promising for the selected variables.

Keywords – Multilayer perceptron neural network, backpropagation, stepwise, electrical load forecasting.

Resumo – Previsão de demanda de cargas é importante para o planejamento da geração e distribuição de energia elétrica. Este trabalho apresenta um modelo híbrido para previsão de cargas a curto prazo usando o método *stepwise* para a seleção de variáveis com auxílio da regressão linear múltipla e da rede neural *Perceptron* multicamadas com o algoritmo da retropropagação para ajuste dos pesos sinápticos. Os resultados das análises se mostraram bastantes satisfatórios para as variáveis selecionadas.

Palavras-chave – Rede neural *Perceptron* multicamadas, algoritmo retropropagação, *stepwise*, previsão de cargas elétricas.

1 Introdução

A análise de carga elétrica contribui para a economia de geração, controle de distribuição, planejamento de manutenção, segurança de todo o sistema elétrico, etc. Fazer previsões de demanda de carga antecipa ações preventivas para manter a geração e a distribuição adequadas para cada perfil de consumo de um determinado período Gross and Galiana (1987).

Para a realização da previsão de cargas, existem vários modelos na literatura como a regressão linear simples ou múltipla, ARIMA de Box & Jenkins, o alisamento exponencial, filtro de Kalman, etc. Box and Jenkins (1976), O'Donovan (1983), Gross and Galiana (1987), Mohgran and Rahman (1989),.

Outro modelo bastante usado é a regressão linear múltipla que modela uma equação de regressão, usando medições anteriores, de tal forma que a previsão se aproxime da observação real Larson and Faber (2010).

As redes neurais artificiais *Perceptron* multicamadas com o algoritmo da retropropagação

também apresentam destaque dentre os modelos de previsão, por sua capacidade de aprendizado por meio de um conjunto de observações (conjunto de padrões), obtendo ótimos resultados de previsão. Outra vantagem das redes neurais artificiais é que não é necessária uma modelagem prévia dos dados. Outro ponto a favor das redes neurais é que em sua fase de treinamento (adaptação dos pesos sinápticos) é feito todo em modo *off-line* e com isso o tempo computacional na fase de previsão é baixo Haykin (2001).

Na literatura especializada existem poucos trabalhos que abordam as duas técnicas (rede neural artificial e regressão linear múltipla) de forma híbrida. Dentre eles ressalta-se o trabalho de Ghunem, Assalen and El-Hag (2012) que utiliza a rede neural *Perceptron* com regressão *stepwise* para previsão de óleo transformador. A rede neural *Perceptron* ficou responsável pela modelagem das relações entre os vários parâmetros de óleo transformador e a regressão pelo ajuste do modelo de previsão.

O presente trabalho pretende apresentar um modelo híbrido de previsão de cargas usando o método *stepwise* para seleção de variáveis com o auxílio da regressão linear múltipla e o método dos

mínimos quadrados e a previsão de cargas utilizando a rede neural artificial *Perceptron* multicamadas com o algoritmo da retropropagação. A proposta aqui apresentada constitui-se em uma nova aplicação para o caso da previsão de carga elétrica.

2 Métodos Aplicado ao Problema de Previsão

2.1. Regressão Linear Múltipla, Método dos Mínimos Quadrados

Antes de falar da regressão linear múltipla, é necessário se definir a correlação linear.

Dados dois conjuntos de n observações, X chamado de conjunto das variáveis de entrada, ou variáveis independentes, ou variáveis explicativas e Y chamado conjunto das variáveis dependentes ou variáveis de saída, a correlação linear r é a medida do quanto as variáveis de entrada do conjunto X influencia na saída (ou resposta) no conjunto Y . A correlação linear pode ser calculada usando a equação (1) Larson and Faber (2010).

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}} \quad (1)$$

Em que $-1 \leq r \leq 1$ e é interpretado da seguinte forma Larson and Faber (2010):

- r próximo de 1 - Correlação positiva forte
- r próximo de -1 - Correlação negativa forte
- r próximo de 0 - Não há correlação linear

Se a correlação linear for forte, seja ela positiva ou negativa, é possível determinar uma equação que melhor se aproxime do conjunto de saída e consequentemente fazer previsão para valores futuros. Esta equação chama-se equação de regressão linear e tem a forma da equação (2) Larson and Faber (2010).

$$\hat{y} = m_0 + m_1x_1 + m_2x_2 + \dots + m_kx_k \quad (2)$$

sendo:

x_i : entrada ou variável independente;

\hat{y} : saída prevista ou variável dependente.

Para se obter uma equação de regressão para um dado caso particular é necessário determinar m_0, m_1, \dots, m_k de tal forma que o erro ou resíduo (diferença entre os valores reais e previstos) seja mínimo para isso é usado o método dos mínimos quadrados. Ruggiero and Lopes (1996). As equações descritas em (3) são usadas neste método.

$$\begin{aligned} A &= X^T X \\ B &= X^T Y \\ m &= A^{-1} B \end{aligned} \quad (3)$$

Cuja solução só é possível desde que $X^T X$ seja não singular.

2.2 Método Stepwise

No conjunto de variáveis independentes pode haver variáveis que pouco influenciam o conjunto de variáveis dependentes (saída). O método *stepwise* é usado para selecionar quais variáveis mais influenciam o conjunto de saída podendo, assim, diminuir o número de variáveis a compor a equação de regressão.

O método é feito de forma iterativa, adicionando (passo *forward*) e removendo variáveis (passo *backward*), a partir de um critério de seleção, um dos critérios de seleção mais usados é o teste F, mas também pode ser feito com o coeficiente correlação linear múltipla, erro quadrático total, critério de informação de Akaike Hoking (1976). Neste trabalho o critério de seleção de variáveis adotado foi o coeficiente de correlação linear, por ser um método bastante simples e de fácil implementação computacional. Os procedimentos realizados no método *stepwise* são:

1. Escolhe-se a variável x_k que possui o maior coeficiente de correlação para entrar no modelo.
2. Uma variável x_i entra no modelo, se o coeficiente de correlação for maior que o anterior, x_i permanece no modelo, caso contrário x_i sai do modelo.
3. x_i sai do modelo e se o coeficiente de correlação for menor que o anterior, x_i fica no modelo, caso contrário, x_i permanece fora do modelo. Este passo é repetido até que não tenha mais x_i para sair do modelo. Terminada esta etapa retorna-se ao passo 2 e este passo continua até que não tenham mais variáveis para entrar no modelo.

A figura (1) apresenta um fluxograma que descreve o método *stepwise*.

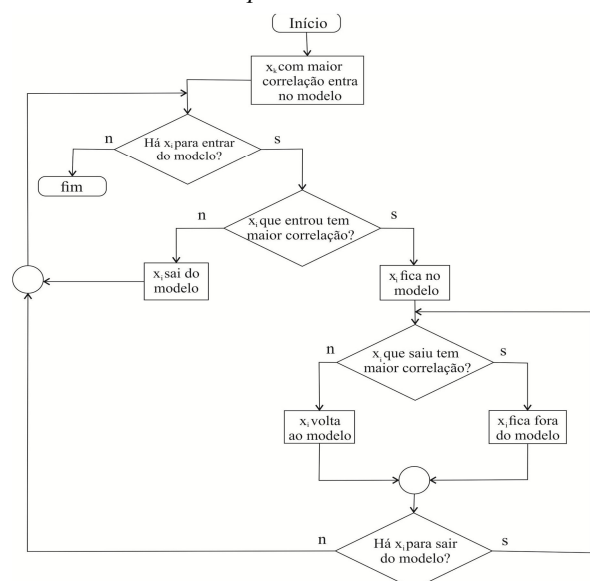


Figura 1: Fluxograma do método *stepwise*.

2.3 Rede Neural Artificial Perceptron Multicamadas

A rede neural *Perceptron* multicamadas é uma importante classe de redes neurais. Ela é do tipo *feedforward* composta por vários neurônios do tipo McCulloch-Pitts dispostos em camadas Haykin (2001), Simpson (1989), Widrow and Lehr (1990). A figura (2) ilustra a arquitetura de uma rede *Perceptron* multicamadas.

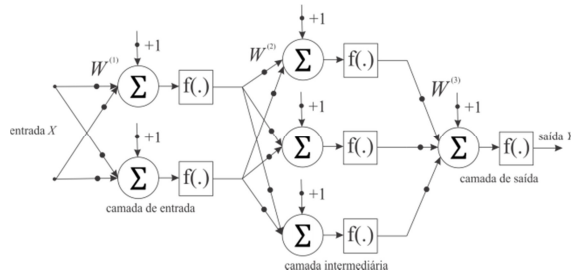


Figura 2: Esquema da rede *Perceptron* Multicamadas.

2.4 Algoritmo de retropropagação

Uma rede neural pode ser treinada com a finalidade de se ajustar seus pesos sinápticos (conjunto W) a fim de produzir uma saída prevista que se aproxime da saída real. Para isso é necessário adotar um algoritmo de adaptação dos pesos. Um dos mais usados é o algoritmo da retropropagação.

Este algoritmo usa o método do gradiente descendente para a correção dos pesos sinápticos, reduzindo o erro produzido pela rede de modo que seja mínimo Haykin (2001), Widrow and Lehr (1990). A seguir os passos que descrevem o *Perceptron* multicamadas com o algoritmo da retropropagação.

1. Um padrão X é apresentado e o sinal propagado através da rede. A saída linear de cada neurônio é dada pela equação (4) Haykin (2001):

$$v = \sum_{i=1}^m w_i x_i + b \quad (4)$$

no qual:

w_i : i -ésimo peso sináptico;
 x_i : i -ésima entrada;
 b : peso correspondente ao “bias”;
 m : número total de entradas.

A saída não linear de cada neurônio é representada pela equação (5) Haykin (2001):

$$y_j(n) = \varphi_j(v_j(n)) \quad (5)$$

sendo:

$y_j(n)$: saída do j -ésimo neurônio na iteração n ;
 φ_j : função não linear.

2. O erro da rede neural na camada de saída da n -ésima iteração é dado pela equação (6) Haykin (2001).

$$e_j(n) = d_j(n) - y_j(n) \quad (6)$$

sendo:

d_j : j -ésima saída desejada;

y_j : j -ésima saída calculada pela rede neural.

3. É calculado o gradiente local referente à última camada através do erro gerado na camada de saída e da derivada do erro através da equação (7) Haykin (2001).

$$\delta_j(n) = e_j(n) \varphi'_j(v_j(n)) \quad (7)$$

A função de não linearidade φ escolhida deve ser diferenciável.

4. Propagação do erro: O gradiente local de cada neurônio das camadas anteriores será calculada através da equação (8) Haykin (2001):

$$\delta_j(n) = \varphi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \quad (8)$$

no qual:

j : índice do neurônio da camada atual;
 k : índice do neurônio da camada imediatamente posterior.

5. Ajuste dos pesos sinápticos: Uma vez calculado cada gradiente local o ajuste dos pesos sinápticos é dado pela equação (9) Haykin (2001):

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_j(n) \quad (9)$$

sendo:

η : taxa de aprendizagem da rede.

A adaptação dos pesos é realizada pela equação (10) Haykin (2001).

$$w_{ji}^{k+1}(n) = w_{ji}^k(n) + \Delta w_{ji}(n) \quad (10)$$

Para cada padrão apresentado à rede é medido o erro instantâneo na equação (11) Haykin (2001).

$$\mathcal{E}(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (11)$$

sendo:

C : conjunto de todos os neurônios da camada de saída.

O algoritmo continua até que todos os padrões sejam apresentados à rede, formando uma época. A convergência do método se dá quando o erro médio dos erros instantâneos atingir uma tolerância pré-estabelecida conforme a equação (12) Haykin (2001).

$$\mathcal{E}_{med} = \frac{1}{N} \sum_{n=1}^N \mathcal{E}(n) \quad (12)$$

no qual:

N : número total de padrões apresentados à rede.

2.5 Critérios para o Perceptron Multicamadas

Antes do processo do *Perceptron* multicamadas as variáveis de entrada oriundas do passo *stepwise* foram modificadas com a finalidade de se melhorar desempenho do modelo.

A variável hora (h) foi binarizada, com valores 1 e -1 e as variáveis de carga foram normalizadas usando o critério dado pela equação (13):

$$\text{carga normalizada} = \frac{\text{carga}}{\max(\text{carga})} \quad (13)$$

A função de ativação da rede escolhida foi a função sigmoide, mostrada na equação (14)

$$f(v) = \frac{1}{1 + e^{-\lambda v}} \quad (14)$$

3 Rede Neural Híbrida

Os experimentos foram realizados usando dados de uma companhia de energia elétrica no ano de 1990. Destes dados foram utilizados, para a seleção de variáveis e a fase de aprendizado da rede, o período de 01/07/1990 a 31/07/1990 e as previsões a curto prazo foram para 01/08/1990 – 24 horas e no período de 01/08/1990 e 02/08/1990 – 48 horas.

Para a seleção de variáveis foi utilizado o método *stepwise* como o auxílio da regressão linear múltipla e o método dos mínimos quadrados. Foram escolhidas variáveis com maior correlação linear e introduzidas no perceptron multicamadas com o algoritmo de retropropagação.

As variáveis disponíveis para as análises foram dados históricos de cargas elétricas, ano, mês, dia da semana e hora referentes à carga elétrica, temperaturas mínima, máxima e média, dados atípicos (feriado), dentre outras.

Na etapa *stepwise* foram selecionadas as variáveis com maior correlação, que foram as variáveis de cargas elétricas e descartadas as variáveis que pouco influenciaram o conjunto de saída, dentre elas as variáveis de temperatura, dia da semana e tipo de dia, para isso foram desprezadas as casas decimais do coeficiente de correlação e dentre os conjuntos de variáveis que apresentaram um ótimo coeficiente de correlação destacaram-se no modelo de previsão:

$$X_{MLP} = [h^{bin} \quad L_n(h) \quad L_n(h-1)]^T \quad (15)$$

e

$$X_{MLP} = [h^{bin} \quad L_n(h-1)]^T \quad (16)$$

Com o conjunto de saídas desejadas

$$Y_{MLP} = [L_n(h+1)]^T \quad (17)$$

no qual:

X_{MLP} : conjunto de variáveis de entrada para o *Perceptron* multicamadas;

h^{bin} : hora binarizada em 5 bits;

$L_n(h)$: carga elétrica normalizada referente a hora (h);

$L_n(h-1)$: carga elétrica normalizada referente a hora ($h-1$);

$L_n(h+1)$: carga elétrica normalizada referente a hora ($h+1$).

4 Resultados

Para avaliar os resultados das análises foram adotados o erro médio percentual absoluto (MAPE) e o erro máximo da previsão que são descritos, pelas equações (18) e (19).

$$MAPE = \frac{1}{n} \sum_{h=1}^{nh} \frac{L(h) - \underline{L}(h)}{L(h)} \times 100 \quad (18)$$

e

$$\text{Erro máximo (\%)} = \max \left(\frac{L(h) - \underline{L}(h)}{L(h)} \right) \times 100 \quad (19)$$

sendo:

$L_n(h)$: carga elétrica referente a hora (h);

$\underline{L}(h)$: carga prevista na hora h ;

nh : número total de horas.

O conjunto de dados de entrada da rede neural artificial representado por:

$$X_{MLP} = [h^{bin} \quad L_n(h) \quad L_n(h-1)]^T$$

apresentou os resultados mostrados na Tabela 1, nas quais, as curvas de previsão demanda de carga elétrica obtidas, tanto para 24 horas quanto para 48 horas, podem ser visualizadas nas Figuras 3 e 4, respectivamente.

Tabela 1. MAPE (%), Erro Máximo (%) RNA Híbrida com as variáveis de entrada hora (h) carga ($h-1$), carga (h).

Correlação	Período	MAPE (%)	Erro Máximo (%)
98%	24 horas	1,32%	4,52%
	48 horas	1,19%	4,52%

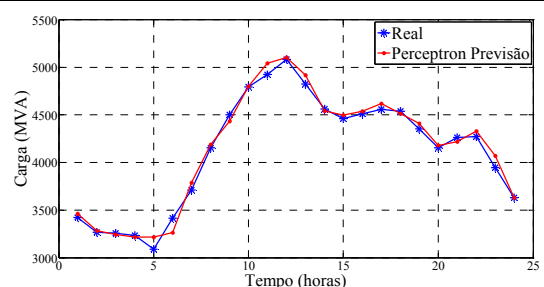


Figura 3: Representação gráfica da previsão realizada com as variáveis hora (h), carga ($h-1$) e carga (h) – 24 horas.

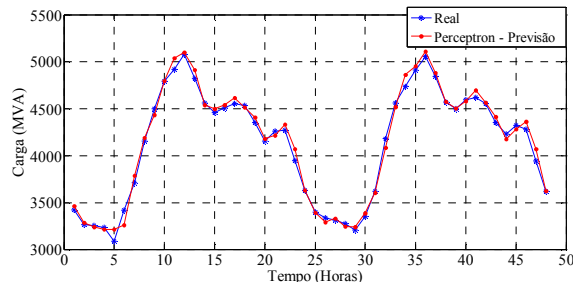


Figura 4: Representação gráfica da previsão realizada com as variáveis hora (h), carga (h-1) e carga (h) - 48 horas.

O conjunto de entrada para o treinamento da rede neural artificial dado por

$$X_{MLP} = [h^{bin} \quad L_n(h-1)]^T$$

forneceu os resultados ilustrados na Tabela 2. As Figuras 5 e 6, respectivamente, mostram as curvas das cargas elétricas reais e obtidas pela rede neural híbrida.

Tabela 2. MAPE (%), Erro Máximo (%) RNA Híbrida com as variáveis de entrada hora (h), carga (h).

Correlação	Período	MAPE (%)	Erro Máximo (%)
97%	24 horas	1,35%	4,62%
	48 horas	1,31%	4,62%

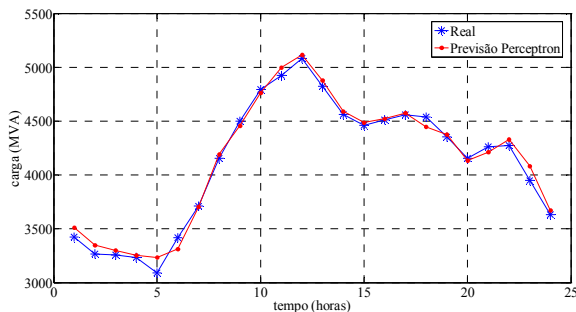


Figura 5: Representação gráfica da previsão realizada com as variáveis hora (h) e carga (h)- 24 horas.

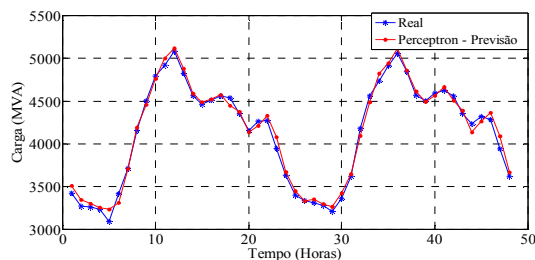


Figura 6: Representação gráfica da previsão realizada com as variáveis hora (h) e carga (h)- 48 horas.

O conjunto de variáveis representado por:

$$X_{MLP} = [L_n(h) \quad L_n(h-1)]^T \quad (20)$$

apresentou um bom valor do coeficiente de correlação, porém, este conjunto de dados não apresentou bom desempenho para o *Perceptron* multicamadas.

Diversas arquiteturas foram testadas com esse conjunto e em todas elas a rede estagnou em pontos de mínimos locais e, assim não convergindo para resultados satisfatórios de previsão. Isso provavelmente se deve ao fato do *Perceptron* multicamadas fazer melhores previsões com uma diversidade de informações como descrito em Haykin (2001), o que não ocorre no conjunto de dados dada pela equação (20).

Foram realizadas análises com variáveis que pouco contribuíram para o aumento do coeficiente de correlação, por exemplo, as variáveis de temperatura. Em todos os casos analisados não se obteve bons desempenhos na rede neural *Perceptron* multicamadas e mesmo usando diversas arquiteturas de rede não convergiram para a tolerância estabelecida, estacionando em mínimos locais.

5 Conclusão

O critério de seleção *stepwise* usando a regressão linear com o auxílio do método dos mínimos quadrados é um processo relativamente simples e com baixo custo computacional agilizando assim, a seleção de variáveis.

Uma das dificuldades do *Perceptron* multicamadas é a determinação de uma arquitetura de rede que melhor se enquadre ao problema apresentado. Quanto mais variáveis o conjunto de padrões tiver maiores serão as possibilidades de arquitetura e consequentemente, mais testes a realizar, por este motivo, a redução de variáveis se torna interessante.

Os resultados obtidos foram satisfatórios, pois dentro de um conjunto de oito variáveis o processo *stepwise* selecionou três variáveis (no caso do primeiro resultado) ou duas variáveis (no caso do segundo resultado), com isso o conjunto de padrões ficou o mais reduzido possível, dentro dos critérios do *stepwise* (maior correlação linear) e do *perceptron* multicamadas (conteúdo diverso).

Para melhorar o desempenho do *Perceptron*, o próximo passo da pesquisa é realizar outras análises usando outro algoritmo de adaptação dos pesos sinápticos, o algoritmo de Levenberg-Maquardt.

6 Referências

Box, G. E. P. and Jenkins, G. M (1976). Times series analysis: forecasting and control, San Francisco.

- Ghunem, R. A., Assaleh, K. and El-Hag, A. H. (2012). Artificial neural networks with stepwise regression for predicting transformer oil furan content, *Dielectrics and Electrical Insulation*, IEEE Transactions on , vol.19, no.2, pp.414-420.
- Gross, G. and Galiana F.D. (1987). Short Term Load Forecasting, *Proceedings of the IEEE*, Vol. 75, pp. 1558-1573. DOI: [10.1109/PROC.1987.13927](https://doi.org/10.1109/PROC.1987.13927)
- Haykin, S. (2001) *Redes Neurais Princípios e Prática*. 2^a. ed, Bookman Porto Alegre-RS.
- Hocking, R. R. (1976). The Analysis and Selection of Variables in Linear Regression. *Biometrics*, v. 32, n. 1, p. 1-49, Washington. DOI: [10.2307/2529336](https://doi.org/10.2307/2529336)
- Larson, R. and Faber, B. (2010). *Estatística Aplicada*, 4^a. ed. Pearson Hall, São Paulo - SP.
- Mohgram, I. and Rahman, S. (1989). Analysis and Evaluation of Five Short term Load Forecasting Techniques, *IEEE Transactions on Power Systems*, Vol. 4, No. 4, pp. 1484-1491. DOI: [10.1109/59.41700](https://doi.org/10.1109/59.41700)
- O'Donovan, T. M. (1983). *Short Term Forecasting: An Introduction to the Box-Jenkins Approach*, John Wiley & Sons, New York.
- Ruggiero, M. A. G. and Lopes, V. L. de R. (1996). *Cálculo Numérico: aspectos teóricos e computacionais*, 2^a. ed. Pearson Makron Books, São Paulo - SP.
- Simpson, P. K. Artificial (1989). *Neural Systems: Foundations, Paradigms, Applications, and Implementations*, Pergamon Press, New York,
- Widrow, B. and Lehr, M. A. (1990). 30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation , *Proceedings of the IEEE*, Vol. 78, No. 9, pp. 1415-1442. DOI: [10.1109/5.58323](https://doi.org/10.1109/5.58323)