

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

MODELOS PARA PREVISÃO DE CARGA A CURTO PRAZO ATRAVÉS DE REDES
NEURAS ARTIFICIAIS COM TREINAMENTO BASEADO NA
TEORIA DA INFORMAÇÃO

WESIN RIBEIRO ALVES

DM 32/2011

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2011

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

WESIN RIBEIRO ALVES

MODELOS PARA PREVISÃO DE CARGA A CURTO PRAZO ATRAVÉS DE REDES
NEURAS ARTIFICIAIS COM TREINAMENTO BASEADO NA
TEORIA DA INFOMAÇÃO

DM 32/2011

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2011

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

WESIN RIBEIRO ALVES

MODELOS PARA PREVISÃO DE CARGA A CURTO PRAZO ATRAVÉS DE REDES
NEURAS ARTIFICIAIS COM TREINAMENTO BASEADO NA
TEORIA DA INFORMAÇÃO

Dissertação submetida à Banca
Examinadora do Programa de Pós-
Graduação em Engenharia Elétrica da
UFPA para a obtenção do Grau de
Mestre em Engenharia Elétrica

UFPA / ITEC / PPGE
Campus Universitário do Guamá
Belém-Pará-Brasil
2011

A474m Alves, Wesin Ribeiro

Modelos para previsão de carga a curto prazo através de redes neurais artificiais com treinamento baseado na teoria da informação /Wesin Ribeiro Alves; orientadora, Adriana Rosa Garcez CastroMaria Emília de Lima Tostes.-2011.

Dissertação (Mestrado) – Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-graduação em Engenharia Elétrica, Belém, 2011.

1. Sistemas de energia elétrica – processamento de dados. 2. Carga e distribuição elétrica. 3. Redes neurais (computação). I. orientador. II. título.

CDD 22. ed. 621.3191

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

MODELOS PARA PREVISÃO DE CARGA A CURTO PRAZO ATRAVÉS DE REDES
NEURAS ARTIFICIAIS COM TREINAMENTO BASEADO NA
TEORIA DA INFORMAÇÃO

AUTOR: WESIN RIBEIRO ALVES

DISSERTAÇÃO DE MESTRADO SUBMETIDA À AVALIAÇÃO DA BANCA
EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-
GRADUAÇÃO EM ENGENHARIA ELÉTRICA DA UNIVERSIDADE FEDERAL DO
PARÁ E JULGADA ADEQUADA PARA OBTENÇÃO DO GRAU DE MESTRE EM
ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA

APROVADA EM 4/11 /2011

BANCA EXAMINADORA:

Prof. Dra. Adriana Rosa Garcez Castro
(ORIENTADORA – UFPA)

Prof. Dra. Maria Emília Tostes
(MEMBRO – UFPA)

Prof. Dr. André Cavalcante do Nascimento
(MEMBRO – IFPA/PA)

VISTO:

Prof. Dr. Marcus Vinícius Alves Nunes
(COORDENADOR DO PPGEE/ITEC/UFPA)

DEDICATÓRIA

Dedico esta vitória aos meus queridos pais, Antônio e Domingas, à minha irmã Susan, à minha namorada Evelyn, e a todos meus amigos que acreditaram e oraram por mim.

Wesin Ribeiro Alves

AGRADECIMENTOS

Em primeiro lugar, agradeço ao meu bom Deus que iluminou a estrada para realizar esse sonho e caminha comigo lado a lado nessa jornada.

Aos meus pais Antônio e Domingas, pois devo a minha vida a eles. Tudo que tenho e tudo que sou é graças aos meus pais.

A professora Dr^a. Adriana Castro, por ter me orientado ao longo deste trabalho e por ter contribuído na correção e ajudado sobretudo para conclusão do mesmo.

Ao professor Dr. Aldebaro Klautau Jr. que se prontificou a dar apoio quando bem precisávamos.

Aos professores que aceitaram participar da banca examinadora

Ao CNPq pelo apoio financeiro.

Wesin Ribeiro Alves

EPÍGRAFE

Deus prova a sua existência através das
entrelinhas da vida.

Wesin Ribeiro Alves

SUMÁRIO

LISTA DE ACRÔNIMOS	XII
NOTAÇÃO MATEMÁTICA	XIII
LISTA DE FIGURAS.....	XV
LISTA DE TABELAS.....	XVI
RESUMO.....	XVII
ABSTRACT	XVIII
1. Introdução	1
1.1 Motivação	1
1.2 Objetivos.....	2
1.3 Organização do trabalho	2
2. Redes Neurais Artificiais.....	4
2.1 Introdução.....	4
2.2 O neurônio artificial.....	4
2.3 Redes Perceptron Multicamadas	6
2.4 Aprendizagem das Redes Neurais	8
2.4.1 Aprendizado baseado no Erro Médio Quadrático	9
2.4.2 Aprendizado baseado na Teoria da Informação	12
2.4.2.1 Entropia	13
2.4.2.2 Janela de Parzen e Estimação da Entropia.....	14
2.4.2.3 Critério da Entropia para Aprendizado de MLP.....	16
2.4.2.4 Correntropia.....	19
3. Previsão de Cargas Elétricas	21
3.1 Introdução.....	21
3.2 A carga elétrica.....	21
3.3 Métodos de previsão de carga	24
3.4 Os diversos horizontes de previsão de carga	25
3.5 Métodos tradicionais de previsão de cargas elétricas.....	26
3.5.1 Método da Regressão Linear ou Múltipla	27
3.5.2 Alisamento Exponencial.....	27
3.5.3 Séries Temporal de Box&Jinks	28
3.5.4 Método de Espaço de Estado e Filtro de Kalman.....	31
3.6 Redes neurais em previsão de carga – Estado da arte	32
4. Propostas de Modelos para Previsão de Carga a Curto Prazo utilizando RNA com Treinamento baseado na Teoria da Informação	34
4.1 Introdução.....	34

4.2	Obtenção e tratamento dos dados	35
4.3	Ambiente computacional	37
4.4	Estrutura da RNA	37
4.5	Modelos de Previsão de Carga através de RNAs com treinamento baseado na Teoria da Informação.....	38
4.5.1	Modelo predictor baseado no critério da Entropia.....	38
4.5.2	Modelo predictor baseado no critério da Correntropia.....	42
4.6	Modelo predictor baseado no MSE.....	47
4.7	Análise comparativa dos modelos predictores propostos	49
5.	Conclusões	53
	Referências Bibliográficas	55
	Apêndice A	60
	Apêndice B	61

LISTA DE ACRÔNIMOS

ANFIS	Adaptive Neuro Fuzzy Inference System
AR	Auto Regressive
ARMA	Auto Regressive Moving Average
ARIMA	Auto Regressive Integrated Moving Average
CIM	Correntropy Induced Metric
DNN	Dinamic Neural Network
fdp	Função Densidade de Probabilidade
ITL	Information Theory Learning
MA	Moving Average
MCC	Maximization Correntropy Criteria
MEE	Minimization Entropy Error
MLP	Multi Layer Perceptron
MSE	Mean Square Error
PI	Potencial da Informação
RNA	Rede Neural Artificial
SOM	Self Organizing Map

NOTAÇÃO MATEMÁTICA

w_{km}	Peso que se encontra na camada k recebendo o sinal do neurônio m
θ_k	Bias da camada k
v_k	Sinal de saída da junção aditiva
$\varphi(\cdot)$	Função de ativação e vetor de coeficientes estatísticos
y_k	Saída do neurônio da camada k
s_j	Saída do neurônio j da ultima camada oculta
β_{kj}	Peso que liga o neurônio da última camada escondida j ao neurônio k da saída
$g(\cdot)$	Função de ativação da camada de saída
$e_j(n)$	Representa o sinal de erro na saída do neurônio j , para a iteração n .
$d_j(n)$	Representa a resposta desejada para o neurônio j .
$y_j(n)$	Representa o sinal funcional que aparece na saída do neurônio j .
$\varepsilon(n)$	Valor instantâneo da energia total do erro
ε_{med}	Função custo ou energia média do erro quadrático
$\Delta w_{ji(n)}$	Fator de correção do peso
$w_{ji(n+1)}$	Peso sináptico da saída do neurônio i à entrada do neurônio j iteração $n+1$
δ_j	Gradiente local induzido
N	Taxa de aprendizagem
H_s	Entropia de Shannon
$H_{R\alpha}$	Família de Entropia de Renyi
H_{R2}	Entropia quadratic de Renyi
$f_Y(z)$	Função densidade de probabilidade da v.a Y
$f'_Y(z)$	Estimativa da função densidade de probabilidade da v.a Y
G, K	Função Kernel
σ	Abertura da janela de Parzen e desvio padrão
$V(y)$	Potencial da Informação
F_{ij}	Força de Informação

$g(x_i, w)$	Mapeador de entrada/saída em relação a entrada x_i e pesos w
$D_g(q; p)$	Distância Csiszar
$V_\sigma(X, Y)$	Correntropia entre as variáveis aleatórias escalares X e Y
$V'_\sigma(X, Y)$	Estimativa da correntropia entre as variáveis aleatórias escalares X e Y
$y(t)$	Carga elétrica no instante t
a_n	Coefficiente de regressão
q_n	Variáveis exógenas
$a(t)$	Variável aleatória com média zero e variância constante
$\varepsilon(t), w(t)$	Ruído branco gaussiano
ϕ_i	Coefficiente de auto-regressão
θ_i	Coefficiente de médias móveis
$A(t)$	Matriz de transição de estados
$B(t)$	Matriz de correlação
$S(t)$	Matriz de covariância

LISTA DE FIGURAS

Figura 2.1 - Neurônio Artificial.....	4
Figura 2.2 - Funções Sigmóides	6
Figura 2.3 - Rede Neural do tipo MLP	7
Figura 2.4 - Ilustração do impacto causado pelo valor de σ na estimação da fdp com Janelas de Parzen.	15
Figura 2.5 - Regiões de CIM(X,0) no espaço de amostra 2D (tamanho do kernel é igual a 1).	20
Figura 3.1 - Série temporal estacionária.....	24
Figura 3.2 - Série temporal não estacionária	25
Figura 4.1 - Série temporal dos dados de treino	36
Figura 4.2 - Topologia da MLP utilizada	37
Figura 4.3 - fdp do erro dos modelos previsores para treinamento com $\sigma = 0.3$	39
Figura 4.4 - fdp do erro dos modelos previsores para treinamento com $\sigma = 0.01$	39
Figura 4.5 - fdp do erro dos modelos previsores para treinamento com $\sigma = 0.1$	40
Figura 4.6 - Histograma do erro para dados de treinamento	41
Figura 4.7 - Histograma do erro para dados de validação	41
Figura 4.8 - Resultados do modelo previsor para 24 horas de um dia	42
Figura 4.9 - fdp do erro dos modelos previsores para treinamento com $\sigma = 0.3$	44
Figura 4.10 - fdp do erro dos modelos previsores para treinamento com $\sigma = 0.01$	44
Figura 4.11 - fdp do erro dos modelos previsores para treinamento com $\sigma = 0.1$	45
Figura 4.12 - Histograma do erro para dados de treino	45
Figura 4.13 - Histograma do erro para dados de validação	46
Figura 4.14 - Resultados do modelo previsor para 24 horas de um dia	46
Figura 4.15 - Histograma do erro para dados de treino	48
Figura 4.16 - Histograma do erro para dados de validação	48
Figura 4.17 - Comparação da fdp entre os modelos para dados de treinamento.....	50
Figura 4.18 - Comparação da fdp entre os modelos para dados de validação.....	50
Figura 4.19 - Resultados para 24 horas do dia	51

LISTA DE TABELAS

Tabela 4.1 - Características Estatísticas das séries utilizadas para treino e validação dos modelos.....	36
Tabela 4.2 - Resultado dos treinos da RNA baseada no critério da Entropia.....	38
Tabela 4.3 - Resultado dos treinos da RNA baseada no critério da Correntropia.....	43
Tabela 4.4 - Resultados dos treinos da RNA baseada no critério MSE	47
Tabela 4.5 - Resultados dos modelos previsores desenvolvidos	49

RESUMO

O conhecimento prévio do valor da carga é de extrema importância para o planejamento e operação dos sistemas de energia elétrica. Este trabalho apresenta os resultados de um estudo investigativo da aplicação de Redes Neurais Artificiais do tipo Perceptron Multicamadas com treinamento baseado na Teoria da Informação para o problema de Previsão de Carga a curto prazo. A aprendizagem baseada na Teoria da Informação se concentra na utilização da quantidade de informação (Entropia) para treinamento de uma rede neural artificial. Dois modelos previsores são apresentados sendo que os mesmos foram desenvolvidos a partir de dados reais fornecidos por uma concessionária de energia. Para comparação e verificação da eficiência dos modelos propostos um terceiro modelo foi também desenvolvido utilizando uma rede neural com treinamento baseado no critério clássico do erro médio quadrático. Os resultados alcançados mostraram a eficiência dos sistemas propostos, que obtiveram melhores resultados de previsão quando comparados ao sistema de previsão baseado na rede treinada pelo critério do MSE e aos sistemas previsores já apresentados na literatura.

PALAVRAS-CHAVE: Previsão de Cargas, Redes Neurais Artificiais, Teoria da Informação

ABSTRACT

The previous knowledge of the load value is almighty important to the electric power system planning and operation. This paper presents results of an investigative study of application of Artificial Neural Networks as a Multilayer Perceptron with the training based on Information Theory to the problem of short term load forecasting. The learning based on Information Theory focuses on the use of the amount of information (Entropy) for the training of neural network. Two forecaster models are presented, and that they was developed using real data from an energy utility. To compare and verify the efficiency of the proposed systems, it was also developed a forecasting system using neural network trained based on the traditional criterion of mean square error (MSE). The results has showed the efficiency of proposed systems, which had better results when compared with the forecasting system based on neural network trained by criterion of MSE and with forecasting system already was presented in the literature.

KEYWORDS: Load Forecasting, Artificial Neural Network, Information Theory

Capítulo 1

Introdução

1.1 MOTIVAÇÃO

Deter o conhecimento da demanda de carga elétrica futura de forma precisa é importante para diversas atividades de planejamento e operação em sistemas de potência. A qualidade do controle do sistema de energia e a economia na sua operação são altamente sensíveis a erros de previsão, logo, estimar a demanda de energia futura de forma correta é prerrogativa obrigatória no gerenciamento da produção e distribuição dessa energia.

Os métodos de previsão de carga podem ser divididos em modelos de previsão a curtíssimo, curto, médio e longo prazo. A previsão a curto prazo trabalha com um tempo de previsão variando de poucas horas a até uma semana a frente e é usada, geralmente, para alimentar ferramentas analíticas de sinalização de tarefas, planejamento do controle de tensão, segurança e economia.

As metodologias de previsão de carga evoluíram muito nos últimos anos. Métodos estatísticos como Auto-Regressão e Séries Temporais vêm sendo extensivamente usados para previsão de carga. Outros modelos tais como Box-Jenkins, ARMA, ARIMA, Filtros de Kalman e modelos baseados na Expansão Espectral também vêm sendo utilizados. Todas estas técnicas trabalham bem em condições normais, entretanto deixam a desejar em situações onde ocorrem mudanças inesperadas nos parâmetros do ambiente. Relacionamentos extremamente complicados entre variáveis que podem levar a operações matemáticas complexas na hora da previsão é outro fator negativo destes modelos.

Para tentar suprir as desvantagens das metodologias tradicionais de previsão, sistemas de previsão baseados em técnicas de Inteligência Computacional vêm sendo desenvolvidos. As Redes Neurais Artificiais vêm sendo bastante utilizadas, principalmente devido a sua alta capacidade para modelagem de problemas não lineares.

Apesar de todos os modelos de previsão de carga com resultados satisfatórios já desenvolvidos, a busca por modelos mais precisos ainda é um ponto de interesse de muitas concessionárias de energia. Considerando os modelos baseados em redes neurais, muitos estudos ainda podem ser desenvolvidos com o intuito de se encontrar ótimas estruturas de redes ou abordagens de treinamento convenientes que, possivelmente, possam melhorar a precisão da previsão.

Em geral, para a realização do treinamento (aprendizado) da rede neural, o erro médio quadrático (MSE) é utilizado como medida de desempenho. Entretanto, novos estudos voltados para a aprendizagem baseada na Teoria da Informação vêm surgindo e resultados promissores têm sido apresentados para problemas de previsão de séries temporais.

A aprendizagem baseada na Teoria da Informação se concentra na utilização da quantidade de informação (Entropia) para treinamento da rede neural. Para o treinamento com a entropia em vez de se utilizar uma medida do conteúdo da informação sobre a variância dos erros (como nos treinamentos baseados no MSE), utiliza-se uma medida do conteúdo da informação da distribuição do erro. Com isto espera-se extrair mais informações dos dados, pois a entropia considera todos os momentos da distribuição do erro. Logo, se a entropia mede o conteúdo em informação de uma distribuição, o objetivo do treinamento das redes com entropia passa então a ser a minimização desta entropia, o que levará, durante a fase de treino, ao acúmulo máximo de informação na estrutura dos pesos da rede (parâmetros ajustáveis durante treinamento) e apenas informação residual na distribuição do erro (distribuição do erro com maior frequência de valores próximos de zero – Função de Dirac).

1.2 OBJETIVOS

Considerando os resultados promissores já apresentados na literatura que mostram as vantagens do treinamento baseado na Teoria da Informação quando aplicadas ao problema de séries temporais, este trabalho tem por objetivo apresentar os resultados de um estudo investigativo da aplicação de redes neurais com treinamento baseado na Teoria da Informação ao problema de previsão de carga. Dois modelos para previsão de carga a curto prazo (horizonte de 1 hora) foram desenvolvidos utilizando Redes Neurais Artificiais do tipo Perceptron Multicamadas considerando dados históricos de carga de uma concessionária do setor elétrico. Para verificação da eficiência dos modelos propostos também foi desenvolvido um sistema previsor utilizando uma rede neural treinada baseada no critério clássico do MSE.

1.3 ORGANIZAÇÃO DO TRABALHO

Além deste capítulo introdutório, esta dissertação segue a seguinte estrutura.

O capítulo 2 apresenta uma visão geral sobre Redes Neurais Artificiais, com ênfase nas redes do tipo Perceptron Multicamadas. É apresentada a descrição da Aprendizagem baseada no Backpropagation clássico e a aprendizagem baseada na Teoria da Informação.

O capítulo 3 apresenta os conceitos sobre previsão de cargas elétricas, mostrando como estas podem ser modeladas, fatores que influenciam no seu comportamento, horizontes de previsão e técnicas clássicas utilizadas para previsão. O capítulo termina com um breve estado da arte de Inteligência Computacional aplicada ao problema de previsão de cargas.

O capítulo 4 apresenta os resultados do estudo investigativo dos modelos previsores construídos com redes neurais do tipo MLPs com treinamento baseado na Teoria da Informação. Esses modelos são comparados a um sistema previsor desenvolvido com redes neurais clássicas do tipo MLPs.

O capítulo 5 apresenta as conclusões, contribuições do trabalho e sugestões para trabalhos futuros.

Capítulo 2

Redes Neurais Artificiais

2.1 INTRODUÇÃO

As Redes Neurais Artificiais são modelos computacionais não lineares, inspirados na estrutura e operação do cérebro humano, que procuram reproduzir características humanas, tais como aprendizado, associação, generalização e abstração. Elas podem ser utilizadas para resolver problemas de aproximação de funções, classificação e previsão de séries temporais.

Neste capítulo serão apresentados os principais conceitos sobre Redes Neurais Artificiais, dando destaque para as Redes Perceptron Multicamadas que é a rede utilizada para o desenvolvimento dos sistemas previsores propostos neste trabalho. Será apresentado também uma breve descrição sobre aprendizado de redes neurais, destacando-se o aprendizado baseado na Teoria da Informação.

2.2 O NEURÔNIO ARTIFICIAL

Inspirado nas propriedades fundamentais do neurônio biológico, o neurônio artificial é o elemento base do processamento de uma Rede Neural Artificial (RNA). O neurônio artificial da Figura 2.1 é composto funcionalmente por três elementos: um conjunto de sinapses (caracterizadas por pesos), uma junção somatória e uma função de ativação.

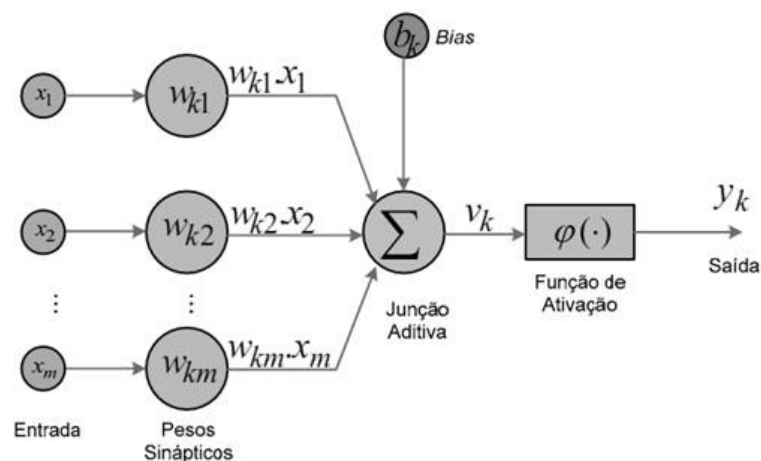


Figura 0.1 - Neurônio Artificial

A conectividade entre os sinais de entrada e o neurônio é realizada através dos pesos, que têm a função de atenuar ou amplificar, através de um fator w_{ij} , o sinal de entrada x_i vindo do ambiente ou de outros neurônios.

A junção somatória, aditiva ou somador realiza a soma ponderada das entradas de acordo com:

$$v_k = \sum_{i=1}^m w_{ki} x_i + b_k \quad 2.1$$

Onde b_k é o *bias* aplicado externamente, que tem o efeito de incrementar ou reduzir o sinal da junção somatória.

O ativador, formado pela função de ativação $\varphi(\cdot)$, recebe o sinal v_k e calcula o nível de estímulo interno do neurônio. Dependendo deste nível, a saída poderá ser ativada ou não. A função de ativação normaliza a amplitude da saída do neurônio no intervalo de $[0 \ 1]$, ou alternativamente $[-1 \ 1]$. A saída do neurônio é então calculada por:

$$y_k = \varphi(v_k) \quad 2.2$$

Na literatura, destacam-se basicamente 3 tipos de funções de ativação: limiar , linear por partes e a sigmóide:

1. *Função limiar*. A saída dessa função apresenta a seguinte forma:

$$\varphi(v_k) = \begin{cases} 1 & \text{Se } v_k \geq 0 \\ 0 & \text{Se } v_k < 0 \end{cases} \quad 2.3$$

2. *Função linear por partes*. A saída dessa função tem o seguinte comportamento:

$$\varphi(v_k) = \begin{cases} 1 & \text{Se } v_k \geq 0.5 \\ v_k & \text{Se } 0.5 > v_k > -0.5 \\ 0 & \text{Se } v_k \leq -0.5 \end{cases} \quad 2.4$$

3. *Função Sigmóide*. Possui a propriedade de ser assintótica e suave. Existem duas formas mais comuns de função sigmóides:

- *Função Logística.* Essa função sigmóide possui um intervalo entre [0 1], conforme a Figura 2.2(a), e é descrita conforme:

$$\varphi(v_k) = \frac{1}{1 + e^{-av_k}} \quad 2.5$$

- *Função Tangente Hiperbólica.* Essa função sigmóide apresenta um intervalo entre [-1 1], conforme a Figura 2.2(b), e é descrita por:

$$\varphi(v_k) = \frac{1 - e^{-av_k}}{1 + e^{-av_k}} \quad 2.6$$

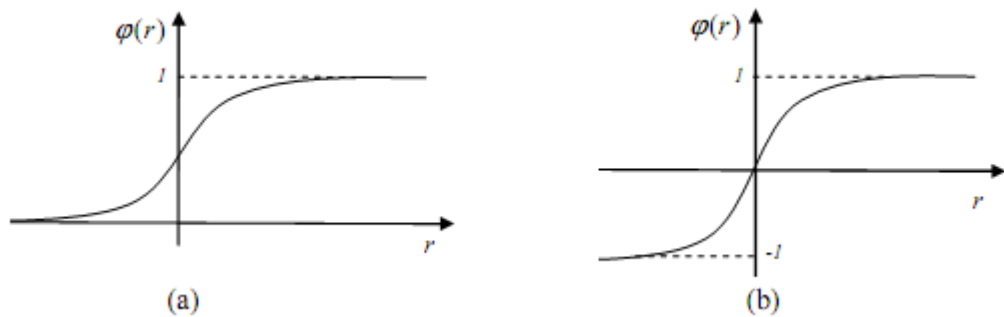


Figura 0.2 - Funções Sigmóides

2.3 REDES PERCEPTRON MULTICAMADAS

As redes de múltiplas camadas alimentadas adiante, também conhecidas como *Perceptron* multicamadas (MLP), são uma importante classe de redes neurais. Uma MLP (Figura 2.3) é constituída por uma camada de entrada, uma ou mais camadas ocultas, e uma camada de saída, sendo que as camadas são compostas de um conjunto de neurônios interconectados. O sinal de entrada se propaga para frente através da rede, camada por camada.

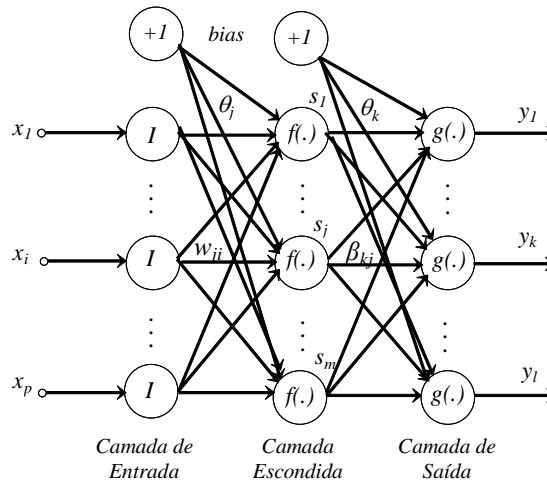


Figura 0.3 - Rede Neural do tipo MLP

A primeira camada oculta é alimentada pela camada de entrada, que também é constituída por neurônios; as saídas resultantes da primeira camada oculta são por sua vez aplicadas à próxima camada oculta e assim sucessivamente até atingir a saída da rede.

Cada neurônio s_j da camada oculta realiza o seguinte cálculo:

$$s_j = f\left(\sum_{i=1}^n w_{ij}x_i + \theta_j\right) \quad 2.7$$

onde x_i é a i -ésima entrada para o neurônio j , w_{ij} é o peso entre o neurônio j e a entrada x_i , θ_j é o bias do neurônio j e $f(\cdot)$ é a função de ativação do neurônio.

Similarmente, considerando a camada de saída, cada neurônio de saída y_k calcula:

$$y_k = g\left(\sum_{j=1}^m \beta_{kj}s_j + \theta_k\right) \quad 2.8$$

onde β_{kj} é o peso que liga o neurônio oculto j da última camada escondida ao neurônio k da saída da rede, θ_k é o bias do k -ésimo neurônio de saída e $g(\cdot)$ é a sua função de ativação.

As MLPs vêm sendo bastante utilizadas na literatura principalmente para problemas de aproximação de função, classificação de padrão e previsão de séries temporais

(Altıparmak, 2009; Upendar, 2010; Sun, 2005). Segundo (Haykin, 2001), uma MLP tem três características distintivas:

1. O uso de uma função de ativação não linear em cada neurônio da rede, sendo que essa função deve ter um coeficiente de não linearidade suave para que seja diferenciável em qualquer ponto.
2. A presença de no mínimo uma camada de neurônios ocultos, que não fazem parte da entrada nem da saída da rede. Esses neurônios ocultos capacitam a rede a aprender tarefas complexas extraíndo progressivamente as características mais significativas dos padrões (vetores) de entrada.
3. A rede exibe um alto grau de conectividade, determinado pelas sinapses da rede. Uma modificação na conectividade da rede requer uma mudança na população das conexões sinápticas ou de seus pesos.

A propriedade mais importante da rede MLP é sua capacidade de ser uma aproximadora universal (Hornik, 1989). Essa capacidade depende exclusivamente da topologia da rede e do número de neurônios que cada camada possui. Entretanto, na literatura nada é conclusivo a respeito da quantidade necessária de neurônios ocultos para encontrar a solução ótima do aproximador universal.

2.4 APRENDIZAGEM DAS REDES NEURAIS

Para que uma rede neural MLP seja útil e consiga desempenhar a função para qual foi designada é necessário passar por uma etapa chamada de aprendizagem.

De acordo com Mendel e MaClaren (Mendel ET AL, 1970): *“Aprendizagem é um processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo contínuo de estímulo pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela forma como a troca de parâmetros se dá”*.

O processo de aprendizagem pode ser classificado em dois paradigmas distintos: aprendizado não-supervisionado e o supervisionado.

O aprendizado não supervisionado é um processo auto-organizável, isto é, ele ocorre em um ambiente onde a saída da rede não depende do meio exterior. Dito de outro modo, a própria rede decide qual é a melhor saída para determinada entrada.

O aprendizado supervisionado depende do ambiente externo, isto é, a saída da rede é comparada a um alvo (valor desejado). Os pesos são ajustados de modo a minimizar o erro

entre a saída da rede e o valor desejado. Dentre os algoritmos usados nesse tipo de aprendizagem destaca-se o algoritmo de Retropropagação do erro.

A idéia principal do algoritmo de retropropagação é realizar o ajuste dos pesos de uma rede neural por meio da propagação inversa do erro da camada de saída. Esse algoritmo é um procedimento sistemático de adaptação dos pesos das redes neurais baseado no método do gradiente descendente (Simpson 1989, Wasserman, 1989).

De modo sucinto, o algoritmo de retropropagação de erro consiste em propagar inversamente os erros da saída da rede neural, e isso é feito em dois fluxos: o fluxo para frente, que se denomina propagação e o fluxo para trás, a retropropagação. Na propagação, o vetor de entrada é aplicado aos neurônios artificiais da rede e seu efeito caminha através da rede, camada após camada. Em seguida, um conjunto de saídas é produzido como a resposta real da rede. Durante o passo de propagação, os pesos sinápticos não sofrem alteração, ou seja, permanecem fixos.

Durante a retropropagação, ocorre o ajuste dos pesos sinápticos de acordo com uma regra de correção de erro, isto é, a resposta real da rede é subtraída de uma resposta desejada (alvo) para produzir um sinal de erro. É esse sinal de erro que é então propagado para trás através da rede, contra a direção das conexões sinápticas. Nesse momento, os pesos sinápticos são alterados para fazer com que a resposta real da rede se mova para mais perto da resposta desejada.

A forma clássica do algoritmo de retropropagação de erros possui o objetivo de minimizar o erro médio quadrático (MSE), que é chamado de função custo, porém essa não é a única função custo que pode ser utilizada. Novos estudos voltados para a aprendizagem baseada na Teoria da Informação vêm surgindo e resultados promissores têm sido apresentados na literatura. A seguir serão apresentados o aprendizado baseado no erro médio quadrático e o aprendizado baseado na Teoria da Informação.

2.4.1 Aprendizado baseado no Erro Médio Quadrático

O sinal de erro de um neurônio j da camada de saída, na iteração n é calculado por:

$$e_j(n) = d_j(n) - y_j(n) \quad 2.9$$

Onde: $e_j(n)$: representa o sinal de erro na saída do neurônio j , para a iteração n .

$d_j(n)$: representa a resposta desejada para o neurônio j .

$y_j(n)$: representa o sinal funcional que aparece na saída do neurônio j .

Define-se o valor instantâneo da **energia do erro** para o neurônio j como sendo $\frac{1}{2}e_j^2(n)$. Ao expandir esse cálculo para todos os neurônios da camada de saída realizando um somatório dos termos, encontra-se o **valor instantâneo da energia total do erro**, conhecido também por erro quadrático $\varepsilon(n)$:

$$\varepsilon(n) = \frac{1}{2} \sum_{j \in c} e_j^2 \quad 2.10$$

Onde: c é o conjunto que inclui todos os neurônios da camada de saída da rede.

Além do valor instantâneo é possível calcular a **energia média do erro quadrático**:

$$\varepsilon_{med} = \frac{1}{N} \sum_{n=1}^N \varepsilon(n) \quad 2.11$$

onde N é o número de padrões apresentados a rede para treinamento.

A energia instantânea do erro $\varepsilon(n)$, e conseqüentemente a energia média do erro ε_{med} , é uma função de todos os parâmetros livres, isto é, pesos sinápticos e níveis de bias da rede. Para um dado conjunto de treinamento, ε_{med} representa a **função de custo ou função objetivo** que é a função utilizada para medir o desempenho de rede durante o processo de aprendizagem.

Durante o treinamento os pesos são ajustados a medida que cada padrão de treino é apresentado à rede. Os ajustes são obtidos de acordo com os respectivos erros calculados para cada padrão. Por conseguinte, a estimativa da alteração real é a média aritmética dessas alterações individuais de peso sobre o conjunto de treinamento, que é baseada na minimização da função custo ε_{med} sobre o conjunto de treinamento inteiro.

A correção $\Delta w_{ji}(n)$ usada para ajuste do peso sináptico $w_{ji}(n+1)$ é proporcional à derivada parcial $\frac{\partial \varepsilon(n)}{\partial w_{ji}}$. Usando a regra da cadeia, tem-se:

$$\frac{\partial \varepsilon(n)}{\partial w_{ji}} = \frac{\partial \varepsilon(n)}{\partial e_j(n)} \cdot \frac{\partial e_j(n)}{\partial y_j(n)} \cdot \frac{\partial y_j(n)}{\partial v_j(n)} \cdot \frac{\partial v_j(n)}{\partial w_{ji}(n)} \quad 2.12$$

Onde: $\partial\epsilon(n)$ é valor instantâneo da energia total do erro na iteração n .

$e_j(n)$: erro na saída do neurônio j , para a iteração n .

$y_j(n)$: valor da saída que aparece no neurônio j , na iteração n .

$v_j(n)$: campo local induzido produzido na entrada da função de ativação do neurônio j .

$w_{ji}(n)$: peso sináptico pertencente a saída do neurônio i à entrada do neurônio j , na iteração n .

A derivada $\frac{\partial\epsilon(n)}{\partial w_{ji}}$ representa um **fator de sensibilidade**, que determina a direção de busca no espaço de pesos para o peso sináptico w_{ji} . Para encontrar o valor de **2.12**, é necessário calcular as derivadas parciais de modo que:

$$\frac{\partial\epsilon(n)}{\partial w_{ji}(n)} = -e_j(n) \cdot \varphi'_j(v_j(n)) \cdot y_i(n) \quad 2.13$$

Então, o fator de correção é obtido:

$$\Delta w_{ji}(n) = -\mu \frac{\partial\epsilon(n)}{\partial w_{ji}(n)} \quad 2.14$$

$$\Delta w_{ji}(n) = -\mu \delta_j(n) y_i(n) \quad 2.15$$

Onde: δ_j é o gradiente local induzido e μ é a taxa de aprendizagem.

De posse do fator de correção é possível encontrar a **regra delta generalizada** que é usada para ajustar o peso w_{ji}^l , da camada l , na iteração $n+1$:

$$w_{ji}^l(n+1) = w_{ji}^l(n) + \mu \delta_j^l(n) y_i^{l-1}(n) \quad 2.16$$

O gradiente local $\delta_j(n)$ poderá assumir duas formas, dependendo da localização do neurônio j : se o neurônio j é um nó de saída, $\delta_j(n)$ é expresso como o produto da derivada $\varphi'_k(v_k(n))$ pelo sinal de erro $e_j(n)$ associado ao neurônio j ; se o neurônio j é um nó oculto, $\delta_j(n)$ é expresso como o produto da derivada associada $\varphi'_k(v_k(n))$ pela soma ponderada dos

δ s calculados para os neurônios na próxima camada oculta ou camada de saída que estão conectados ao neurônio j :

$$\delta_j^{(l)} = \begin{cases} \varphi'_j(v_j^{(L)}(n))e_j^{(L)}(n) & \text{para o neurônio } j \text{ da camada de saída } L \\ \varphi'_j(v_j^{(l)}(n))\sum \delta_k^{(l+1)}(n)w_{kj}^{(l+1)}(n) & \text{para o neurônio } j \text{ da camada oculta } l \end{cases} \quad 2.17$$

Segundo (Jacobs, 1998), o algoritmo da retropropagação tende a convergir lentamente devido a sua natureza estocástica, ou seja, possui uma tendência de procurar os pontos de forma aleatória, em uma espécie de “zigue-zague” em busca da verdadeira direção que leva a um mínimo na superfície de erro, ou em outras palavras, esse algoritmo utiliza uma “estimativa instantânea” conhecida como aproximação estocástica para calcular o gradiente local de cada neurônio. Haykin aponta algumas possíveis causas para esse comportamento, tais como: a superfície do erro, que pode ser razoavelmente plana ou muito curva ao longo de uma dimensão de peso; a direção do vetor gradiente negativo pode não apontar para um mínimo da superfície de erro; a existência de mínimos locais (Haykin, 2001).

Não existe até o momento um algoritmo comprovadamente eficiente para assegurar um treinamento ótimo (Haykin, 2001).

2.4.2 Aprendizado baseado na Teoria da Informação

A aprendizagem baseada na Teoria da Informação se concentra na utilização da quantidade de informação (Entropia) para treinamento da rede neural (Príncipe, 2009). Para o treinamento utilizando a entropia em vez de se utilizar uma medida do conteúdo da informação sobre a variância dos erros, se utiliza uma medida do conteúdo de informação da distribuição do erro.

Com o uso da entropia para treinamento das redes neurais espera-se extrair mais informações dos dados, pois a entropia considera todos os momentos da distribuição do erro. Logo, se a entropia mede o conteúdo em informação de uma distribuição, o objetivo do treinamento das redes com entropia passa então a ser a minimização desta entropia, o que levará, durante a fase de treino, ao acúmulo máximo de informação na estrutura dos pesos da rede (parâmetros ajustáveis durante treinamento) e apenas informação residual na distribuição do erro (distribuição do erro com maior frequência de valores próximos de zero – Função de Dirac).

O aprendizado baseado na Teoria da Informação (ITL) pode ser implementado tanto pela Entropia quanto pela Correntropia . Na prática, o aprendizado baseado na teoria da informação substitui os momentos de segunda ordem: a variância pela entropia; a correlação pela correntropia; e a minimização do erro quadrático (MSE) pela minimização da entropia do erro ou maximização da correntropia.

2.4.2.1 Entropia

A Teoria da Informação diz que quanto menos previsível for um evento, maior será seu valor de informação, ou em outras palavras, o valor da informação é inversamente proporcional a probabilidade da ocorrência de um evento, sendo esse conceito conhecido como Entropia. Embora isso pareça algo abstrato, Shannon (Shannon,1948) traduziu matematicamente a definição de entropia de uma distribuição discreta de probabilidade $P = (p_1, p_2, \dots, p_n)$ como sendo:

$$H_S(P) = \sum_{k=1}^N p_k \log \frac{1}{p_k}, \text{ com } \sum_{k=1}^N p_k = 1 \text{ e } p_k \geq 0 \quad 2.18$$

Shannon deu apenas o primeiro passo sobre os conceitos de entropia, pois alguns anos depois de seu trabalho surgiu uma forma mais genérica de calcular a quantidade de informação de uma mensagem chamada de Entropia de Renyi (Renyi,1976). De fato, esse novo conceito é uma família de entropias $H_{R\alpha}$ que dependem do parâmetro α . Quando $\alpha = 2$, ela se torna a denominada Entropia Quadrática de Renyi:

$$H_{R\alpha} = \frac{1}{1-\alpha} \log \sum_{k=1}^N p_k^\alpha \text{ com } \alpha > 0, \alpha \neq 1 \quad 2.19$$

$$H_{R2} = -\log \sum_{k=1}^N p_k^2 \quad 2.20$$

Existe uma relação bem conhecida entre a entropia de Shannon e Renyi:

$$H_{R\alpha} \geq H_S \geq H_{R\beta} \quad \text{se } \beta > 1 > \alpha > 0 \quad 2.21$$

$$\lim_{\alpha \rightarrow 1} H_{R\alpha} = H_S$$

A entropia quadrática pode ser generalizada para uma variável aleatória contínua Y com função de densidade de probabilidade (fdp) $f_Y(z)$:

$$H_{R2} = -\log \int_{-\infty}^{+\infty} f_Y(z)^2 dz \quad 2.22$$

A entropia de Shannon, como um somatório pesado de logaritmos, não é atrativa do ponto de vista do esforço computacional. A entropia de Renyi, que é o logaritmo da soma de probabilidades, é a maneira menos dispendiosa de se calcular a Entropia, e é por isso que ela foi escolhida para ser aplicada no aprendizado baseado na Teoria da Informação.

2.4.2.2 Janela de Parzen e Estimação da Entropia

A estimativa de uma função densidade de probabilidade de dados de uma amostra constituída de pontos discretos $y_i \in R^M$, $i=1, \dots, N$ em um espaço M-dimensional, pode ser realizada pelo método da Janela de Parzen (Parzen, 1962). Esse método usa um somatório de funções Kernel (função nuclear) para estimar a fdp da distribuição de uma amostra, onde o mais recomendável é o Kernel Gaussiano, pois ele é continuamente diferenciável e provê menos esforço computacional devido a propriedade que diz que a convolução de uma função gaussiana também é uma gaussiana, com covariância igual a soma das covariâncias individuais. Esta técnica usa a função Kernel centrada em cada ponto da amostra – interpreta cada ponto, como sendo localmente descrito por uma fdp de Dirac. O ponto é substituído (aproximado) por um conjunto contínuo cuja densidade de probabilidade é descrita pelo *Kernel*. A expressão da estimativa f'_Y da fdp f_Y de um conjunto de N pontos é encontrada a partir da soma das contribuições individuais do *Kernel* aplicado em cada ponto na forma:

$$f'_Y(z) = \frac{1}{N} \sum_{i=1}^N G(z - y_i, \sigma^2 I) \quad 2.23$$

onde o vetor $y_i \in \mathcal{M}$ são amostras da variável aleatória contínua, $K(\cdot)$ é a função Kernel e σ o tamanho da janela de *Parzen*, sendo o Kernel gaussiano dado por:

$$G(z_k - y_{ik}, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(z_k - y_{ik})^2} \quad 2.24$$

onde $\sigma^2 I$ é a matriz de covariância (aqui assumida com variâncias iguais e independentes em todas as dimensões). A matriz de covariâncias é diagonal com todos os elementos da diagonal principal iguais e os restantes nulos e pode ser representado a partir de uma matriz identidade

I multiplicada por uma constante σ^2 . No fundo tem-se uma fdp centrada no valor do ponto i na sua dimensão k e com variância σ^2 .

O tamanho da Janela de Parzen é definido pelo valor do parâmetro σ , que denota uma estimativa suave para f_Y quando seu valor é alto, caso contrário, denota uma estimativa mais abrupta, conforme pode ser observado na Figura 2.4.

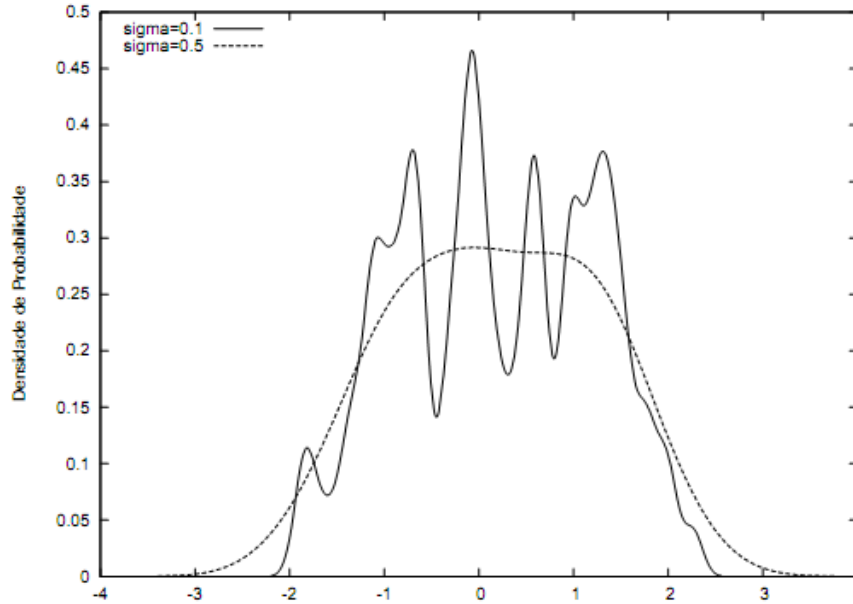


Figura 0.4 - Ilustração do impacto causado pelo valor de σ na estimação da fdp com Janelas de Parzen.

A estimativa da entropia H_{R2} pode ser realizada então aplicando-se o conceito de janelas de Parzen. Substituindo na equação 2.22 a fdp f_y pela sua aproximação obtida em 2.23 tem-se:

$$H_{R2} = -\log \int_{-\infty}^{\infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G(z - y_i, \sigma^2 I) G(z - y_j, \sigma^2 I) dz \quad 2.25$$

ou

$$H_{R2} = -\log \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \int_{-\infty}^{\infty} G(z - y_i, \sigma^2 I) G(z - y_j, \sigma^2 I) dz \quad 2.26$$

levando a :

$$H_{R2} = -\log V(y) \quad 2.27$$

com:

$$V(y) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \int_{-\infty}^{+\infty} G(z - y_i, \sigma^2 I) G(z - y_j, \sigma^2 I) \quad 2.28$$

Aplicando-se a propriedade da convolução de gaussianas tem-se então:

$$V(y) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G(y_i - y_j, 2\sigma^2 I) \quad 2.29$$

ou

$$V(y) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N V_{ij} \quad 2.30$$

A expressão 2.29, que é conhecida na Teoria da Informação como Potencial de Informação (PI) e permite obter, de maneira algoritmicamente tratável a Entropia pelo simples cálculo dos valores das funções gaussianas do vetor de distância entre os pares de pontos de amostras y . É neste ponto que pode ser observada a diferença conceitual entre o critério baseado na teoria da informação e o critério baseado no erro médio quadrático. A minimização da variância (critério do MSE) recai apenas na informação sobre cada ponto individualmente, enquanto que a otimização pelo critério da Entropia exige a consideração da relação mútua entre pares de pontos sendo por este motivo que se consegue a extração de mais informação da distribuição dos pontos y .

Como o logaritmo é uma função monótona e crescente, minimizar a entropia equivale a maximizar o potencial de informação PI.

2.4.2.3 Critério da Entropia para Aprendizado de MLP

O critério da Entropia para aprendizado de uma MLP propõe que o ajuste dos pesos seja realizado através da aplicação de um método de otimização para minimização da equação 2.27 (minimização da entropia da distribuição dos erros).

De 2.29 e 2.30 tem-se que:

$$V_{ij} = G(y_i - y_j, 2\sigma^2 I) \quad 2.31$$

Verifica-se que 2.31 representa uma interação entre as partículas de informação i e j que depende da distância entre elas $d_{ij} = y_i - y_j$

V_{ij} representa de fato uma interação mútua entre partículas e a soma das interações de todas as partículas sobre a partícula i é dada por:

$$V_i = \sum_{j=1}^N V_{ij} \quad 2.32$$

Desta forma cada partícula apresenta uma energia própria V_{ii} e está submetida a interações mútuas V_{ij} . V_{ii} apresenta valor constante dado por:

$$V_{ii} = G(0, 2\sigma^2 I) \quad 2.33$$

A variação $\frac{\partial}{\partial y_i} V_{ij}$ na interação entre as partículas será:

$$\frac{\partial}{\partial y_i} G(y_i - y_j, 2\sigma^2 I) = -G(y_i - y_j, 2\sigma^2 I)(y_i - y_j) \frac{1}{2\sigma^2} = F_{ij} \quad 2.34$$

sendo F_{ij} designado como Força de Informação (FI).

O efeito acumulativo de todas as partículas sobre a partícula i é dado por:

$$F_i = \frac{\partial}{\partial y_i} V(y) = -\frac{1}{N^2 \sigma^2} \sum_{j=1}^N G(y_i - y_j, 2\sigma^2 I)(y_i - y_j) = -\frac{1}{N^2 \sigma^2} \sum_{j=1}^N V_{ij} d_{ij} \quad 2.35$$

Então, para se encontrar os pesos de uma rede MLP que usa o algoritmo de retropropagação para minimizar a Entropia dos erros na saída da rede, aplica-se a derivada em relação ao peso w_k em 2.30 considerando a partícula y como sendo o erro e entre a saída de um neurônio de saída da MLP e a resposta desejada:

$$\frac{\partial}{\partial w_k} V(e(w)) = \sum_{i=1}^N \frac{\partial V}{\partial e_i} \frac{\partial e_i}{\partial w_k} = \sum_{i=1}^N F_i \frac{\partial e_i}{\partial w_k} \quad 2.36$$

A atualização dos pesos é realizada então na direção contrária ao gradiente do PI:

$$\Delta w_k = -\mu \frac{\partial V(e)}{\partial w_k} \quad 2.37$$

onde μ é a taxa de aprendizado.

Fazendo uma analogia com a física, as amostras de dados podem ser consideradas como uma partícula em campo de informação. Existe entre elas uma relação mútua de acordo com leis de interação definidas, sendo que os parâmetros da rede neural são alterados de acordo com a interação entre essas partículas.

Se a entropia das amostras for minimizada, as forças são atrativas (com sinal positivo devido à maximização do potencial) e procuram mover as partículas em direção ao potencial máximo. Se, por outro lado, a entropia das amostras for maximizada, as forças se comportam repulsivamente e o movimento das partículas é de se espalhar para preencherem o espaço.

A manipulação dos pesos w da rede RNA é o mecanismo que faz movimentar as partículas de informação. Por essa razão, é necessário propagar o efeito da força de informação transformando o movimento das partículas no espaço de saída em movimento dos pesos no próprio espaço.

Logo, verifica-se que a força de informação tem o mesmo papel que o erro desempenha no critério MSE. Em vez de haver uma retropropagação do erro para alterar os pesos w , há a retropropagação da força de informação.

É importante salientar que ao aprendizado baseado no critério da Entropia é independente do método de treinamento, ou seja, ela pode ser aplicada não somente ao algoritmo de retropropagação dos erros, mas em qualquer outro algoritmo de treino para otimização de MLPs.

As propriedades da entropia estimada pelo método da Janela de Parzen podem ser encontradas em (Erdogmus, 2002). Eis duas propriedades que são essenciais para compreender a sua aplicação no treino das MLPs:

Propriedade 1: O estimador da entropia não é sensível a média fdp, isto é, a resposta do sistema pode convergir para um conjunto de pesos que originam erros com médias afastadas de zero. Isso pode ser corrigido aplicando um bias à resposta para forçar a erro do treino ser nulo.

Propriedade 2: Se o máximo valor do Kernel $K(\xi)$ ocorre quando $\xi = 0$, então o valor mínimo da entropia é obtido quando todas as amostras são iguais umas as outras, isto

é, $x_1 = x_2 = \dots = x_n = c$. No treino supervisionado, se todas as amostras de erro forem igual a zero, a função custo deve atingir o resultado ótimo global.

O Apêndice A apresenta a formulação e o algoritmo desenvolvido para treinamento das MLPs baseado no critério da Entropia e que foi utilizado para os treinamentos dos modelos de Previsão de carga apresentados neste trabalho.

2.4.2.4 Correntropia

De uma forma geral, a correntropia é uma medida de similaridade entre duas variáveis aleatórias escalares X e Y definida pela expressão:

$$V_{\sigma}(X, Y) = E[k_{\sigma}(X - Y)] \quad 2.39$$

onde k é a função de kernel Gaussiana e $E[k_{\sigma}(X - Y)]$ é o valor médio dessa Gaussiana.

Da mesma maneira que a janela de Parzen pode ser usada para estimação da entropia, ela também pode ser usada para a correntropia a fim de se obter uma estimativa para treino de MLPs:

$$V'_{\sigma}(X, Y) = \frac{1}{N} \sum_{i=1}^N k_{\sigma}(x_i - y_i) \quad 2.40$$

Como pode ser visto em 2.29 e 2.40, tanto entropia quanto correntropia usam uma função com kernel Gaussiano, porém, a correntropia tem apenas um simples somatório de termos e é dividida somente por N, portanto, é necessário menos esforço computacional para ser implementada. A maneira de como obter 2.39 de 2.40, propriedades e uma definição completa sobre Correntropia é mostrado em (Liu, 2007).

Tem sido provado em (Liu, 2007) que a correntropia é positiva e limitada a: $0 < V(X, Y) \leq 1/\sqrt{2\pi\sigma}$, e atinge seu valor máximo se e somente se $X=Y$. A correntropia também induz uma métrica no espaço de amostra pela função $CIM(X, Y) = (k(0) - V(X, Y))^{1/2}$, denominada Métrica de Correntropia Induzida. Esta medida pode ser dividida em três diferentes regiões, chamadas de distâncias L ou normas L. Quando o erro da saída do sistema é próximo de zero, CIM é equivalente a norma L2; quando o erro tende a aumentar CIM torna-se norma L1; quando o erro é muito grande CIM torna-se norma L0, a medida

saturo e torna-se muito insensível a erros elevados (Liu, 2007). Essas regiões podem ser vistas na Figura 2.5.

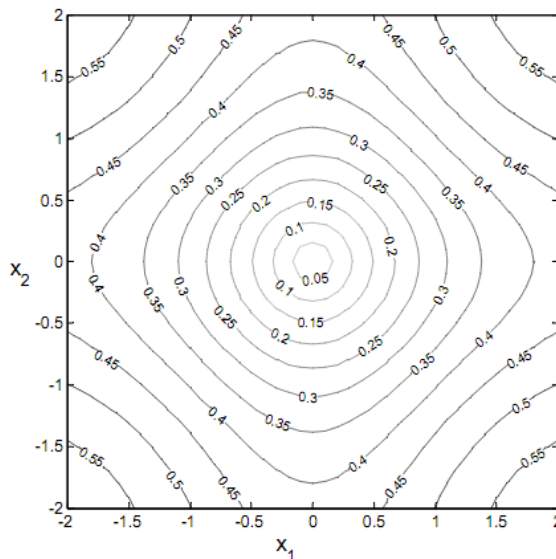


Figura 0.5 - Regiões de $CIM(X,0)$ no espaço de amostra 2D (tamanho do kernel é igual a 1).

Esta propriedade destaca a robustez do CIM e a importância da largura de banda do kernel, ou em outras palavras, a correntropia se comporta como o critério clássico do erro médio quadrático (MSE) quando o tamanho do kernel é grande e aumenta a região Euclidiana, caso contrário, ela tende para uma pequena região Euclidiana.

Portanto, a fim de usar a correntropia para treinar mapeadores, é necessário maximizar a correntropia da distribuição de erros. Então, a saída do sistema se aproxima da resposta desejada no senso da métrica CIM. Assim, a correntropia pode ser usada como uma nova função custo, com a vantagem sobre MSE e entropia de ser um critério local de similaridade, muito útil para casos de média sem zero, sem gaussianas com altos deslocamentos, e por requerer menos esforço computacional.

O Apêndice B apresenta a formulação e o algoritmo desenvolvido para treinamento das MLPs baseado no critério da Correntropia e que foi utilizado para os treinamentos dos modelos de Previsão de carga apresentados neste trabalho.

Capítulo 3

Previsão de Cargas Elétricas

3.1 INTRODUÇÃO

A previsão de carga é uma das principais funções na operação de um sistema de energia elétrica. A qualidade do controle do sistema de energia e a economia na sua operação são altamente sensíveis a erros de previsão. Estimar a demanda de energia futura de forma correta é prerrogativa obrigatória no gerenciamento da produção e distribuição da energia.

Os métodos de previsão de carga podem ser divididos em modelos de previsão a curtíssimo, curto, médio e longo prazo. As metodologias de previsão de carga evoluíram muito nos últimos anos. Métodos estatísticos como Auto-Regressão e Séries Temporais vêm sendo extensivamente usados para previsão de carga. Outros modelos tais como Box-Jenkins, ARMA, ARIMA, Filtros de Kalman e modelos baseados na Expansão Espectral também vêm sendo utilizados.

Este capítulo apresenta uma visão geral sobre o problema de previsão de cargas elétricas. Serão apresentados os métodos tradicionais para previsão de carga. O capítulo termina com a apresentação do estado da arte em Inteligência Computacional (Redes Neurais Artificiais) aplicada ao problema de previsão de cargas.

3.2 A CARGA ELÉTRICA

Quando milhares de aparelhos domésticos ou as máquinas em uma indústria são ligados, a demanda de carga elétrica apresenta uma tendência que pode ser prevista estatisticamente. Essa tendência é influenciada por fatores que interferem na forma de modelar a carga.

A carga elétrica de um sistema de energia é formada por diferentes unidades consumidoras. Uma boa parte é consumida em atividades industriais. De outro lado, tem-se uma parte que é consumida pelo usuário doméstico, por meio, por exemplo, de chuveiros elétricos, iluminação, e aparelhos eletrodomésticos ou eletrônicos.

A demanda de carga elétrica no setor industrial é estimável e previsível, e depende do nível da produção. Entretanto, alguns fatores como máquinas quebradas e até mesmo greve de trabalhadores, que são eventos inesperados, podem afetar de maneira sensível o nível de

consumo. Com relação ao usuário doméstico, é mais difícil de prever um padrão de consumo, pois cada um possui uma maneira diferente de consumir energia. Fatores sociais e comportamentais podem afetar o nível da carga doméstica, como grandes eventos, feriados, e até mesmo programas de TV. Além desses, o clima pode ser determinante no nível de consumo, pois pode aumentar ou diminuir o uso de chuveiros elétricos ou condicionadores de ar, dependendo da temperatura.

Durante o processo de previsão de cargas uma análise a priori, portanto, deve ser realizada para que possa ser verificada a influência dos diversos fatores sobre o comportamento da carga. Os principais fatores que podem influenciar no trabalho de previsão de carga são:

- **Fatores climáticos:** ocorrem a curto prazo podendo citar como exemplo a temperatura, velocidade do vento, umidade relativa do ar, neblina, chuva, temporal com relâmpago e trovão, e neve. A temperatura é um fator muito importante, pois sofre influência dos demais fatores citados, e, com isso, afeta diretamente o consumo de energia elétrica delimitando por exemplo o uso do aparelho de ar-condicionado, aquecedor e chuveiro elétrico (Swarup, 2002; Hippert, 2001; Murto, 1998). Portanto, as condições meteorológicas são responsáveis por variações significativas nos padrões da carga.
- **Fatores econômicos e demográficos:** são fatores que acontecem a longo prazo e que possuem importância fundamental na evolução da previsão de carga. Exemplos desses fatores são: área demográfica, que pode ser rural, urbana, ou industrial, crescimento industrial e econômico, variação na agricultura, alteração no preço da eletricidade, demanda de carga, etc (Murto,1998). Logo, o ambiente econômico e demográfico em que o sistema de potência opera tem um impacto claro sobre a demanda de carga elétrica.
- **Fatores de tempo:** são caracterizados pelos comportamentos cíclicos e sazonais. No primeiro, observa-se o comportamento de carga em dias especiais, tais como sábado e domingo, que sofrem uma redução significativa na demanda, e na segunda-feira e sexta-feira, onde há apenas uma pequena redução da carga. No segundo comportamento, observa-se a variação da carga em determinados períodos do ano, por exemplo: férias, feriados, início do ano escolar, natal, ano novo, carnaval, etc. (Murto, 1998);

- **Fatores esporádicos:** esses fatores causam perturbações na carga. Citando alguns exemplos têm-se os eventos esportivos, programas de televisão, greve em indústrias, etc. (Murto, 1998). Um sistema de potência é continuamente exposto a distúrbios esporádicos devido ao fato que a carga do sistema é composta de um grande número de demandas individuais. Esses fatores podem causar não apenas variações pequenas, mas também grandes variações no consumo de eletricidade.

Existe ainda outro tipo de classificação com relação aos fatores que influenciam a demanda de carga elétrica, a saber:

- Origem
- Tipo
- Duração
- Efeito

Quanto a Origem, pode-se se subdividir em: sociocultural, onde jogos da copa do mundo e programas de televisão são alguns exemplos; econômicos, onde o plano real e tarifação de energia são bons exemplos; climática, no qual a temperatura, umidade e luminosidade podem ser citadas como exemplo; e por fim, a política legal, onde o horário de verão, propaganda eleitoral gratuita são alguns dos exemplos.

Quanto ao Tipo, pode se subdividir ainda em: localizado, cujo apenas parte da curva de carga é afetada; e disperso, onde toda a curva de carga é afetada, ou a sua maior parte.

Quanto a Duração, pode-se se subdividir em: efetivo, provocado por mudanças sazonais; esporádico, onde propaganda eleitoral gratuita e entrada de frente fria são alguns dos exemplos; periódico, provocado, por exemplo, pelo horário de verão; e único, ocasionado por algum desligamento.

Quanto ao Efeito, pode-se se subdividir em: transportador, cujo efeito apenas modula a curva de carga, como, por exemplo, a propaganda eleitoral; redutor, cujo efeito reduz a energia total, como ocorre durante as greves de um grande consumidor; e por fim, o incrementador, que aumenta o consumo, como, por exemplo, algum plano econômico.

Então existem diversos fatores com diversos tipos de classificação que podem influenciar a carga e determinar o nível de consumo de energia, entretanto, a escolha das variáveis que mais sensibilizam a carga é que vai determinar o sucesso da previsão dentro de uma taxa de erro aceitável.

3.3 MÉTODOS DE PREVISÃO DE CARGA

Em (O’ Donovan, 1983) são sugeridas três classes onde se enquadram os métodos de previsão de carga:

1 **Métodos Intuitivos:** Fazem parte dessa classe métodos que não usam um modelo matemático formal para descrever o comportamento da carga.

2 **Métodos Autônomos:** Fazem parte dessa classe métodos que utilizam modelos baseados apenas na extrapolação de observações históricas de carga. Esse método compreende uma análise da evolução da carga em relação ao tempo podendo ser classificada como sendo horizontal, sazonal e tendenciosa, de acordo com determinados padrões de comportamento. No padrão horizontal, os dados flutuam em torno de um determinado valor médio de carga constante no tempo. No padrão sazonal, a carga sofre influência de fatores temporais, tais como: estações climáticas, feriados, etc. Para o padrão tendencioso, os dados flutuam de acordo com uma curva de tendência (média variável no tempo). Portanto, em cada um desses padrões é levado em consideração as observações passadas da carga.

3 **Métodos Causais:** Fazem parte dessa classe modelos onde a carga exhibe uma relação de causa e efeito com relação a uma ou mais variáveis independentes, que irão descrever seu comportamento. Exemplos de variáveis utilizadas para justificar a tendência de crescimento da carga são: índices socioeconômicos, crescimento populacional, etc.

Ainda segundo (O’Donovan, 1983), visando descrever a evolução da carga com relação ao tempo existem duas categorias dessa série temporal, que são utilizadas para determinar a escolha do método de previsão adequado:

1 **Estacionária:** Uma série temporal é definida como estacionária quando os dados da carga permanecem dentro de um intervalo em torno de uma média constante no tempo. A Figura 3.1 apresenta exemplos de dados que flutuam em torno de uma média constante no tempo, ou seja, uma série temporal estacionária.

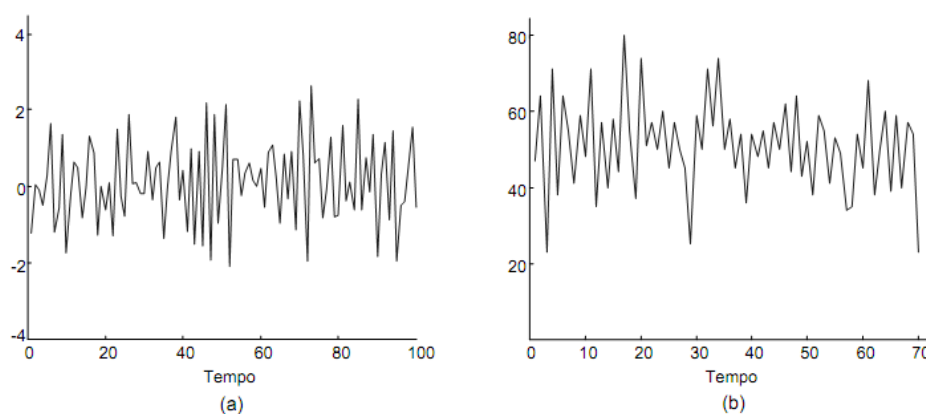


Figura 0.1 - Série temporal estacionária

2 Não-estacionária: A série temporal é definida como não-estacionária se os dados da carga variam com o passar do tempo, ou em outras palavras, se o comportamento histórico dos dados apresenta algum padrão tendencioso, cíclico, sazonal ou aleatório. A Figura 3.2, apresenta dois exemplos de séries não estacionárias. Na Figura 3.2 (a) tem-se um comportamento de uma série não estacionária com padrão sazonal sobreposto semelhante ao padrão horizontal e na Figura 3.2 (b) pode se observar o comportamento de uma série não estacionária com o padrão tendencioso para cima.

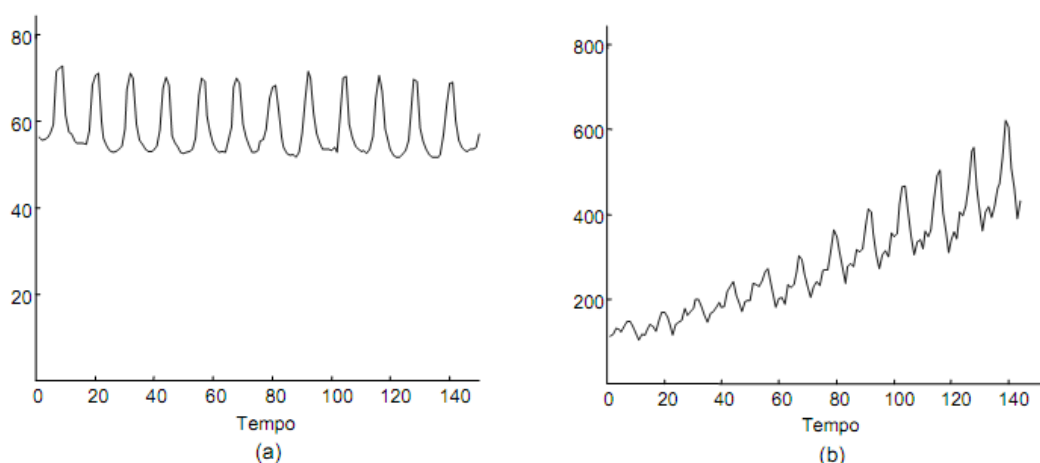


Figura 0.2 - Série temporal não estacionária

3.4 OS DIVERSOS HORIZONTES DE PREVISÃO DE CARGA

A previsão de demanda de cargas elétricas pode ocorrer tanto para fornecimento quanto para consumo e pode estar contida entre diversos horizontes de previsão tais como horas, dias, meses ou anos. Sendo assim, pode-se classificar a previsão de carga de acordo com o período de análise. De acordo com (Murto 1998) existem quatro tipos de previsão de carga:

1. **Previsão de carga a longo prazo:** ideal para operação de planejamento em sistemas de potência pois engloba um período entre 10 a 20 anos. Os dados de cargas elétricas diárias ou semanais são desconsiderados. Também, pode ser aplicados na previsão de equipamentos elétricos, preços de eletricidade, preços de fontes alternativas de energia.
2. **Previsão de carga a médio prazo:** esse tipo de previsão é recomendado para o planejamento de suprimento de combustível, programas de manutenção, e

planejamento de produção a médio prazo. Ela engloba um período de poucas semanas a poucos anos.

3. **Previsão de carga a curto prazo:** esse tipo de previsão é ideal para planejamento de operações econômicas e de segurança diária dos sistema de potência. O intervalo de previsão varia no período de meia hora até poucas horas, ou até mesmo um dia a uma semana, ou seja, a previsão de carga a curto prazo trabalha com a estimativa de carga para cada hora do dia, pico de carga diário, ou geração de energia diária ou semanal.
4. **Previsão de carga a curtíssimo prazo:** esse tipo de previsão é usado essencialmente para previsão em tempo real, ou seja, pode ser utilizado para planejamento de produção *online*, análise de contingência *online*, e detectar condições perigosas de operação. O período de previsão varia entre poucos segundos a alguns minutos (no máximo 15 minutos).

A medida que o horizonte de previsão decresce, a necessidade de fazer uma análise dos dados passados aumenta, ou seja, na medida que o período de previsão decresce, as variações cíclicas do comportamento da carga torna-se mais importante para a tarefa de previsão. Entretanto, quando o patamar de curto e curtíssimo prazo é atingido, outro fator se torna fundamental: a adaptabilidade da previsão. Neste caso deve ser levado em conta os impactos climáticos que atuam sobre a carga, monitorando-se a variação da carga observada em relação a prevista, afim de efetuar algum reparo no sistema.

Não são apenas os níveis de precisão e detalhe que aumentam para sistema que levam em consideração os horizontes de curto e curtíssimo prazo, mas também, a robustez do modelo adotado para previsão, haja vista a necessidade de recuperação imediata em situações inesperadas.

3.5 MÉTODOS TRADICIONAIS DE PREVISÃO DE CARGAS ELÉTRICAS

De acordo com (Gross, 1987; Bunn, 1985) existem várias técnicas de previsão de cargas elétricas e elas podem ser categorizadas em duas classes fundamentais que são os modelos não dinâmicos e dinâmicos. Nos modelos não dinâmicos a carga é expressa como uma série temporal discreta de valores previstos para cada hora durante um período. Em contrapartida, nos modelos dinâmicos a carga é expressa não apenas como uma função do tempo de determinado dia, mas também do seu comportamento passado, por isso a previsão feita de uma hora atual depende das horas anteriores.

3.5.1 Método da Regressão Linear ou Múltipla

A carga pode assumir um comportamento com padrão estocástico em um sistema de potência. Esse modelo descreve esse comportamento. Ele faz uma estimativa da carga a partir da relação da carga padrão e algumas variáveis exógenas, variáveis externas ao modelo, que podem ser climáticas e não climáticas. Temperatura, umidade do ar e luminosidade são alguns exemplos de variáveis climáticas, enquanto que número de consumidores, atividade econômica e renda são exemplos de variáveis não climáticas. A carga pode ser expressa como (Mogham, 1989; Swarup, 2002; Murto, 1998):

$$y(t) = a_0 + \sum_{i=1}^n a_i q_i(t) + a(t) \quad 3.1$$

Sendo $y(t)$ a carga elétrica no instante t , $a_0, a_1, a_2, \dots, a_n$ os coeficientes de regressão, q_1, q_2, \dots, q_n as variáveis exógenas e $a(t)$ a variável aleatória com média zero e variância constante.

As variáveis exógenas são identificadas através de uma análise de correlação junto a carga. Quando o número de variáveis identificadas é muito grande faz-se necessário o uso de um método chamado de STEPWISE (Draper, 1981) para selecionar as variáveis mais significativas. Os coeficientes de regressão são identificados, por sua vez, através do método dos mínimos quadrados. Em seguida, são realizados testes estatísticos tais como o t-Student para determinar a significância dos coeficientes a serem utilizados nessa técnica. Embora esse método seja de fácil implementação, o conhecimento prévio sobre as variáveis exógenas tem limitado o seu uso. Uma descrição mais completa sobre esse método pode ser encontrada nos seguintes trabalhos (Thompson, 1976; Heinemann, 1966; Cargil, 1971; Asbury, 1975).

3.5.2 Alisamento Exponencial

Nesse método a carga pode ser modelada, em um determinado tempo t , utilizando uma função de ajuste (Mogham, 1989). A equação da modelagem da carga e a função de ajuste são dadas por:

$$y(t) = \varphi(t)^T f(t) + \varepsilon(t) \quad 3.2$$

$$f(t) = Lf(t-1) \quad 3.3$$

onde $\varphi(t)$ é um vetor de coeficientes que é estimado usando-se a técnica de mínimos quadrados ponderados para os N intervalos mais recentes. “ T ” indica que o vetor de coeficientes sofre uma transposição. O termo $\varepsilon(t)$ é uma variável aleatória conhecida como ruído branco. Nesse modelo, o vetor dos coeficientes e o valor da carga podem ser calculados recursivamente através de:

$$\varphi'(N+1) = L^T \cdot \varphi'(N) + F^{-1} \cdot f(0) \cdot [y(N+1) - y'(N)] \quad 3.4$$

$$y'(N+1+l) = f^T(l) \cdot \varphi'(N+1) \quad 3.5$$

sendo F e L matrizes construídas em função do histórico da carga. Mais detalhes sobre essa técnica e sua aplicação no ramo da previsão de cargas elétricas a curto prazo pode ser vista em (Lisejen, 1971; Christiaanse, 1971; Gupta, 1972).

3.5.3 Séries Temporal de Box&Jinks

É o método mais utilizado na previsão de carga a curto prazo, e também vem sendo bastante utilizado em previsões a longo e médio prazo (Gupta, 1971; Huang, 2003; Wei, 2009). O comportamento temporal da carga pode ser transformado em uma série temporal estacionária (invariante no tempo), e esse é o princípio básico desse modelo. Para tal tarefa, é necessário um filtro linear que tem como entrada séries aleatórias $a(t)$ e produz a carga transformada em série temporal estacionária $y(t)$. As entradas aleatórias do filtro possuem médias zero e variância constante $\sigma_a^2(t)$, conhecido como ruído branco.

Os modelos estacionários Ruído Branco, *Auto Regressive (AR)*, *Moving Average (MA)* e *Auto Regressive Moving Average (ARMA)* são as classes que pertencem a esse método (O’Donova, 1983; Moghram, 1989; Murto, 1998). Além dessas classes, existem ainda os modelos não estacionários tais como *Walk Aleatório* e o *Auto Regressive Integrated Moving Average*. A seguir será apresentada uma breve descrição sobre cada modelo.

- **Modelo Ruído Branco:** esse modelo é definido pela equação:

$$y(t) = \theta_0 + a(t) \quad 3.6$$

em que $y(t)$ é a carga elétrica; θ_0 é o termo constante; $a(t)$ representa a série aleatória.

- **Modelo *Auto Regressive* ($AR(p)$):** A equação que define esse modelo é:

$$y(t) = \theta_0 + \sum_{i=1}^p \phi_i y(t-i) + a(t) \quad 3.7$$

sendo que os termos acrescentados nessa equação em comparação com a equação 3.6 são os coeficientes de auto-regressivos ϕ_i e a ordem do modelo p .

- **Modelo *Moving Average* ($MA(q)$):** a equação abaixo define esse tipo de modelo:

$$y(t) = \theta_0 + a(t) - \sum_{i=1}^q \theta_i a(t-i) \quad 3.8$$

em que θ_i representa os coeficientes de médias móveis e q representa a ordem do modelo.

- **Modelo *Auto Regressive Moving Average* ($ARMA(p,q)$):** esse modelo é composto por ambos os termos *auto regressive* e *moving average*, ou seja, é expresso pela seguinte equação:

$$y(t) = \theta_0 + \sum_{i=1}^p \phi_i y(t-i) + a(t) - \sum_{j=1}^q \theta_j a(t-j) \quad 3.9$$

- **Modelo *Walk Aleatório*:** esse modelo é expresso pela seguinte equação:

$$y(t) = y(t-1) + \theta_0 + a(t) \quad 3.10$$

Observa-se que a partir das primeiras diferenças regulares desse modelo, forma-se um modelo de Ruído Branco. O último valor apresentado na série temporal é exatamente a previsão para todos os períodos de tempo futuro nesse modelo.

- **Modelo *Auto Regressive Integrated Moving Average (ARIMA)*:** Observa-se que a partir das primeiras diferenças regulares desse modelo, formam-se os modelos ARMA tais como $AR(1)$, $AR(2)$, $MA(1)$ e $MA(2)$ (O'Donovan, 1983).

Existem três possibilidades para esse modelo. A primeira delas é conhecida como *Autoregressive Integrated ARI(1,1)*, na qual é expressa por:

$$y(t) = \theta_0 + (1 + \phi_1)y(t-1) - \phi_1 y(t-2) + a(t) \quad 3.11$$

O modelo $ARI(2,1)$ é encontrado da mesma maneira. Observa-se que as primeiras diferenças regulares dos modelos $ARI(1,1)$ e $ARI(2,1)$ formam, respectivamente, os modelos $AR(1)$ e $AR(2)$.

A segunda possibilidade é o modelo *Integrated Moving Average IMA(1,1)* definido como sendo:

$$y(t) = \theta_0 + y(t-1) - \theta_1 a(t-1) + a(t) \quad 3.12$$

O modelo $IMA(2,1)$ é encontrado de forma similar. Observa-se que as primeiras diferenças regulares dos modelos $IMA(1,1)$ e $IMA(1,2)$ formam, respectivamente, os modelos $MA(1)$ e $MA(2)$.

Por fim, a terceira possibilidade é o modelo $ARIMA(1,1,1)$, que é descrito por:

$$y(t) = \theta_0 + (1 + \phi_1)y(t-1) - \phi_1 y(t-2) + a(t) - \theta_1 a(t-1) \quad 3.13$$

De forma geral, os modelos $ARIMA(p, I, q)$ são encontrados da mesma maneira, isso quer dizer também que as suas primeiras diferenças regulares formam os modelos $ARMA(p, q)$.

Uma descrição mais completa sobre esse método e algumas aplicações em previsão de demanda de carga elétrica a curto prazo podem ser encontrados em (Vemuri, 1981; Hagan, 1978).

3.5.4 Método de Espaço de Estado e Filtro de Kalman

Esse método modela a carga no tempo t como uma variável de estado de um sistema de equações lineares, conhecidas como equação de estado e equação de medida:

$$y(t+1) = A(t)y(t) + w(t) \quad 3.14$$

$$z(t) = B(t)y(t) + v(t) \quad 3.15$$

Sendo $y(t)$ a carga elétrica, $A(t)$ a matriz de transição de estados, $w(t)$ o ruído branco com matriz de covariância $Q(t)$, $z(t)$ o vetor de medição no tempo t , $B(t)$ a matriz de correlação entre $y(t)$ e $z(t)$ e $v(t)$ o erro medido (ruído branco com matriz de covariância $S(t)$).

Para calcular a estimativa *a posteriori* do processo são utilizadas as expressões a seguir (Caines, 1987):

$$y(t/t) = y(t/t-1) + k(t)[z(t) - B(t)y(t/t-1)] \quad 3.16$$

$$k(t) = p(t/t-1)B(t)^T [B(t)p(t/t-1)B(t)^T - S(t)^{-1}] \quad 3.17$$

$$p(t/t) = [1 - k(t)B(t)]p(t/t-1) \quad 3.18$$

A estimativa *a priori* é encontrada através de:

$$y(t+1/t) = A(t)y(t/t) \quad 3.19$$

$$p(t+1/t) = A(t)p(t/t)Q(t)^T + Q(t) \quad 3.20$$

O termo $k(t)$ é conhecido como ganho do Filtro de Kalman e $p(t/t-1)$ é a matriz de covariância do vetor de erro entre o valor atual e a estimativa do processo.

Embora esse método seja muito utilizado para previsão de processos em tempo real, devido a característica de recursividade do filtro, existe muita dificuldade na identificação das

matrizes de covariância $Q(t)$ e $S(t)$. Mais detalhes e alguns trabalhos que utilizam essa técnica podem ser encontrados em (Toyoda, 1970; Abu-El-Magd, 1981; Campo, 1997).

3.6 REDES NEURIAS EM PREVISÃO DE CARGA – ESTADO DA ARTE

As metodologias tradicionais apresentadas na seção anterior, embora tenham evoluído bastante nos últimos anos, apresentam um rendimento inferior quando trabalham em situações onde ocorrem mudanças inesperadas nos parâmetros do ambiente e quando possui variáveis com relacionamentos extremamente complexos.

As técnicas baseadas em Inteligência Computacional vêm sendo implementadas na área de previsão de carga para tentar suprir as desvantagens dos métodos tradicionais como podemos ver em (Amjad, 2006), (Barzamini, 2005), (Castro, 2002), e (Liao, 2006). Dentre elas, destacam-se as Redes Neurais Artificiais devido a sua alta capacidade de modelagem para sistemas não lineares.

Essas ferramentas que auxiliam na tomada de decisão são indispensáveis para que os resultados do sistema predictor de carga sejam ótimos e as perdas mínimas. No contexto das redes neurais, existem vários trabalhos na literatura que abordam o uso dessa técnica para previsão de cargas elétricas.

Em (Yun, 2008) foi feito um estudo de previsão de cargas elétricas a curto prazo abordando o ambiente de preço em tempo real utilizando um modelo híbrido formado por uma rede neural artificial com função de base radial (RBF) e um sistema de inferência neuro-fuzzy adaptativo (ANFIS). Esse modelo primeiramente faz a previsão da carga através da RBF sem considerar o preço da energia, e então, baseado na variação dos preços em tempo real, ele usa o ANFIS para ajustar o resultado de previsão feita anteriormente pela rede neural com base radial, aumentando a precisão do sistema.

A rede neural bayesiana foi desenvolvida em (Ning, 2010) para previsão de cargas elétricas nos dias úteis e finais de semana. Esse modelo utiliza a aprendizagem bayesiana que permite obter os valores mais prováveis dos parâmetros a fim de construir a melhor arquitetura da rede neural. Os dados de treino e teste foram coletados de cargas históricas do sistema de energia da província Guizhou. Os resultados mostraram que a rede neural bayesiana era mais eficiente que os outros modelos.

Já em (Osman, 2009) foi realizado um estudo da técnica de previsão de carga a curto prazo baseada apenas na rede neural artificial que usa não apenas a variável temperatura, mas também, variáveis meteorológicas correlacionadas para treino, validação, e teste, com dados

do Sistema Unificado Egípcio. A análise da correlação entre as variáveis permite determinar as entradas da rede.

Em (Xinbo, 2011), o algoritmo de retropropagação da rede neural usada para previsão de carga num sistema de empréstimo de energia teve seu desempenho melhorado. A adição da constante de momento e da categorização dos dados de entrada com ativação linear foram os meios utilizados para aumentar a precisão desse modelo em comparação a rede neural com retropropagação clássica.

Em (Harum, 2010) foi criado um sistema predictor de cargas usando a técnica de otimização de treino Levenberg-Marquardt utilizada no algoritmo de retropropagação da RNA. A entrada desse modelo era composta de múltiplos atrasos no pico de carga a cada hora permitindo que a informação do movimento padrão da série temporal tivesse sido obtido. O pico de carga horária da Malásia foi o estudo de caso desse modelo.

Em (Chogumaira, 2010) foi apresentado uma previsão de cargas elétricas usando a rede neural dinâmica (DNN), onde os pesos são atualizados dinamicamente. Nesse estudo foi feito uma comparação entre três diferentes arquiteturas aplicadas ao novo modelo dinâmico: MLP, Elman e RNA com função de base radial. Segundo os resultados apresentados, a RNA com função de base radial obteve melhor desempenho que as demais.

Apesar de todos os modelos de previsão de carga com resultados satisfatórios já desenvolvidos, a busca por modelos mais confiáveis ainda é um tema de interesse de muitas concessionárias de energia. Levando em consideração os modelos baseados em redes neurais, ainda há a necessidade de se encontrar ótimas estruturas de redes ou abordagens de treinamento convenientes que, possivelmente, possam melhorar a precisão da previsão.

Capítulo 4

Propostas de Modelos para Previsão de Carga a Curto Prazo utilizando RNA com Treinamento baseado na Teoria da Informação

4.1 INTRODUÇÃO

A previsão de carga a curto prazo, que pode variar de minutos a dias, vem se tornando cada vez mais importante para as empresas, desde a criação dos mercados de energia. O estudo do comportamento de como a carga elétrica irá se apresentar no futuro garante um fornecimento de energia segura, contínua e econômica.

Os modelos baseados nas metodologias tradicionais continuam sendo aplicados ao problema da previsão de carga. Atualmente, entretanto, diversos modelos baseados em Inteligência Computacional vêm surgindo e entre as metodologias de Inteligência Computacional se destacam as Redes Neurais Artificiais. As Redes Neurais Artificiais vêm sendo bastante utilizadas, principalmente devido a sua alta capacidade para modelagem de problemas não-lineares

Apesar de todos os modelos de previsão de carga com resultados satisfatórios já desenvolvidos, a busca por modelos mais confiáveis ainda é de grande interesse de muitas concessionárias de energia. Levando em consideração os modelos baseados em Redes Neurais, ainda há a necessidade de se encontrar ótimas estruturas de redes ou abordagens de treinamento convenientes que, possivelmente, possam melhorar a precisão da previsão.

Considerando então a importância e a necessidade de desenvolvimento de modelos de previsão cada vez mais precisos, neste capítulo serão apresentadas as propostas de modelos de previsão de carga a curto prazo (uma hora adiante) utilizando Redes Neurais Artificiais do tipo MLP com treinamento baseado na Teoria da Informação. Dois modelos foram desenvolvidos considerando dados históricos de carga de uma concessionária do setor elétrico. Para verificação da eficiência dos modelos propostos também foi desenvolvido um sistema predictor utilizando uma Rede Neural treinada baseada no critério clássico do MSE. Os resultados alcançados são favoráveis aos sistemas predictores baseados na Teoria da Informação.

4.2 OBTENÇÃO E TRATAMENTO DOS DADOS

A qualidade da previsão de cargas através das RNAs depende de dois fatores: da escolha das variáveis de entrada mais significativas para o modelo e da disponibilidade do histórico dessas variáveis, tendo em vista que o processo de aprendizagem da rede neural está diretamente ligado a esses dados.

Além da carga, é importante estabelecer quais outras grandezas podem ter influência sobre o modelo previsor. Uma análise qualitativa dos dados também é importante para que um processo de tratamento seja realizado caso seja necessário.

Como já mencionado no Capítulo 3, os fatores que podem influenciar a demanda de carga elétrica são: fatores climáticos, econômicos, temporal e esporádicos. No caso da previsão de carga a curto prazo, a variável que mais exerce influência sobre a curva da carga é a temperatura. As sazonalidades como o dia da semana e as estações do ano também afetam o comportamento da carga. Os jogos da copa, fim de novelas e reality shows, que são fatores aleatórios, também alteram a carga e causam muitos problemas para previsão.

Devido à dificuldade para obtenção de um banco de dados de outras variáveis que poderiam influenciar o comportamento da carga, utilizou-se neste trabalho apenas os valores da carga para os sistemas de previsão através das Redes Neurais propostos. Os dados utilizados para o desenvolvimento dos modelos foram fornecidos pela companhia Centrais Elétricas do Pará (CELPA), sendo que os mesmos foram coletados entre os dias 25/10/2006 e 30/11/2006 em transformadores do bairro do Tapanã na cidade de Belém, e são valores de carga relativos ao ano de 2005, 2006 e 2007.

Para o treinamento da rede, os dados foram divididos em dados de treinamento e dados de validação. Um processo de tratamento foi realizado nos dados para identificar dados faltosos. A base formada pelos dados dos anos de 2005 e 2006, que resultou em um total de 5096 padrões, foi utilizada na fase de treino da RNA, enquanto que a base de dados do ano de 2007, que resultou em um total de 2854 padrões, foi utilizada para validação. A Figura 4.1 apresenta o comportamento da série temporal utilizada para formação dos dados de treinamento da rede neural.

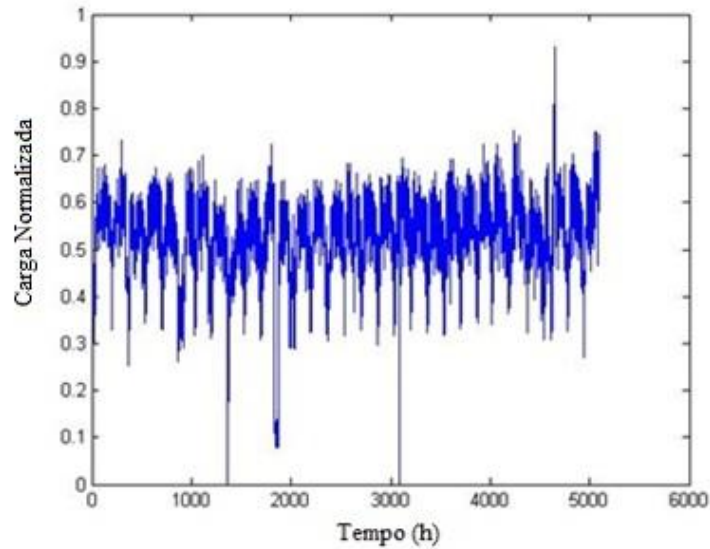


Figura 0.1 - Série temporal dos dados de treino

Toda a base de dados usada no treinamento das RNAs e na validação foi normalizada entre os valores [0 1] utilizando para isto:

$$x_{ij_{norm}} = 2 \frac{(x_{ij} - x_{\min})}{(x_{\max} - x_{\min})} - 1 \quad 4.1$$

onde x_{\min} e x_{\max} correspondem aos valores mínimo e máximo da variável a normalizar e $x_{ij_{norm}}$ é o valor normalizado da variável x_{ij} .

A Tabela 4.1 apresenta as características estatísticas das séries utilizadas para treino e validação dos modelos considerando os dados normalizados.

Tabela 0.1 - Características Estatísticas das séries utilizadas para treino e validação dos modelos.

	Treino	Validação
Média	0.5304	0.577
Variância	0.0086	0.0073
Valor Mínimo	0	0,3
Valor Máximo	0.932	0.905

As Redes Neurais foram treinadas para realizar a previsão de carga uma hora adiante. Após vários testes variando-se o número de entradas da RNA, constatou-se que duas entradas eram suficientes para modelagem dos sistemas, sendo que essas entradas correspondem aos

valores da carga atual e de uma hora passada, isto é, as duas entradas correspondem ao valor da carga no instante t e $t-1$. A saída da rede neural compreende um valor real contendo o valor previsto da carga para uma hora adiante, isto é, $t+1$.

4.3 AMBIENTE COMPUTACIONAL

O software Matlab R2008b foi a ferramenta utilizada para desenvolver, treinar e validar os sistemas previsores de cargas elétricas propostos neste trabalho. A máquina utilizada possuía um processador Intel Pentium Core Duo de 1,86GHz com 2GB de memória RAM.

4.4 ESTRUTURA DA RNA

A estrutura da rede neural MLP escolhida para realizar a previsão de carga com treinamento baseado tanto nos critérios da Teoria da Informação quanto do MSE é composta de 3 camadas: a camada de entrada que é composta por duas variáveis, valor da carga V em t e $t-1$, a camada oculta que é composta de neurônios trabalhando com função de ativação tangente hiperbólica, e a camada de saída trabalhando com função de ativação linear contendo uma variável de saída que é o valor desejado da carga V uma hora adiante ($t+1$). A Figura 4.2 apresenta a topologia da MLP utilizada para desenvolvimento dos modelos de previsão.

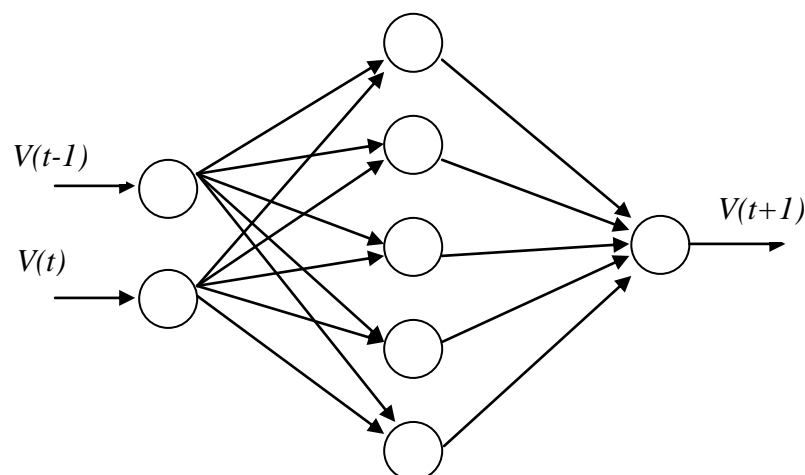


Figura 0.2 - Topologia da MLP utilizada

4.5 MODELOS DE PREVISÃO DE CARGA ATRAVÉS DE RNAs COM TREINAMENTO BASEADO NA TEORIA DA INFORMAÇÃO

4.5.1 MODELO PREVISOR BASEADO NO CRITÉRIO DA ENTROPIA

Para o desenvolvimento do modelo predictor baseado na MLP treinada pelo critério da Entropia é importante destacar que diversas topologias (de acordo com a Figura 4.2) foram treinadas e testadas exaustivamente variando-se o número de neurônios da camada escondida e o valor da abertura da janela de Parzen. A Tabela 4.2 apresenta alguns dos resultados desses diversos treinamentos. Para cada variação do número de neurônios na camada oculta (5, 10, 25 e 45), foram testados valores distintos da variância/sigma (0.3, 0.01 e 0.1) das funções gaussianas das janelas de Parzen. Como critério de parada para os treinamentos adotou-se a variação do valor da entropia ao longo do treinamento: quando este valor já não variava ao longo das iterações, o treinamento era finalizado.

Tabela 0.2 - Resultado dos treinos da RNA baseada no critério da Entropia

Entropia Sigma 0.3	5 neurônios na camada oculta	10 neurônios na camada oculta	25 neurônios na camada oculta	45 neurônios na camada oculta
Nº de iterações	4000	5000	4000	6000
Valor da entropia	-0.1161	-0.1152	-0.1153	-0.1169
Erro Médio (treino)	0.0017	0.0018	0.0019	0.0016
Erro Médio (validação)	0.0015	0.0017	0.0017	0.0014
Entropia Sigma 0.01	5 neurônios na camada oculta	10 neurônios na camada oculta	25 neurônios na camada oculta	45 neurônios na camada oculta
Nº de iterações	1000	1000	500	500
Valor da entropia	-0.93	-0.92	-0.92	-0.89
Erro Médio (treino)	0.0016	0.0016	0.0016	0.0016
Erro Médio (validação)	0.0014	0.0014	0.0014	0.0014
Entropia Sigma 0.1	5 neurônios na camada oculta	10 neurônios na camada oculta	25 neurônios na camada oculta	45 neurônios na camada oculta
Nº de iterações	1000	1000	500	500
Valor da entropia	-0.55	-0.54	-0.55	-0.55
Erro Médio (treino)	0.0015	0.0016	0.0016	0.0016
Erro Médio (validação)	0.0014	0.0014	0.0014	0.0014

Os gráficos da função densidade de probabilidade (fdp) do erro para os modelos apresentados na Tabela 4.2 são apresentados nas Figuras 4.3, 4.4 e 4.5 para cada valor de sigma e variando-se o número de neurônios da camada oculta.

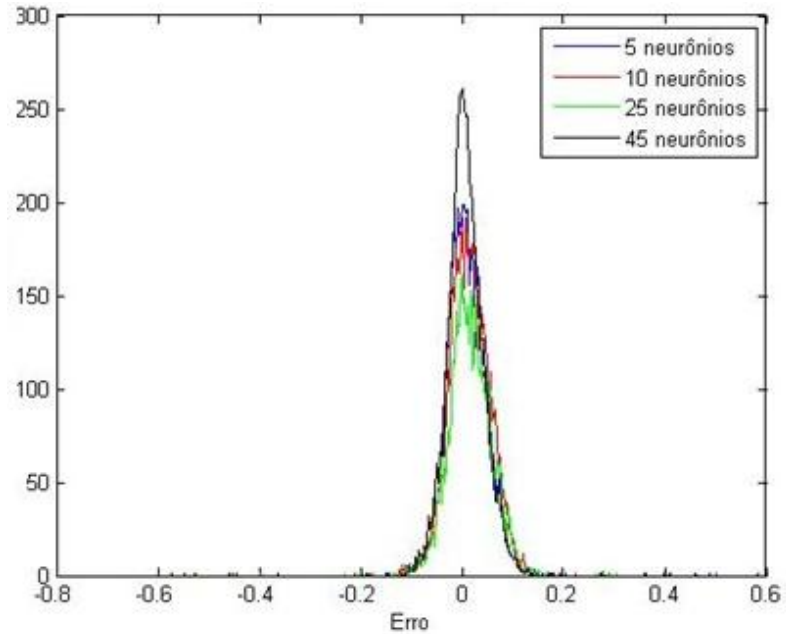


Figura 0.3 - fdp do erro dos modelos previsores para treinamento com $\sigma = 0.3$

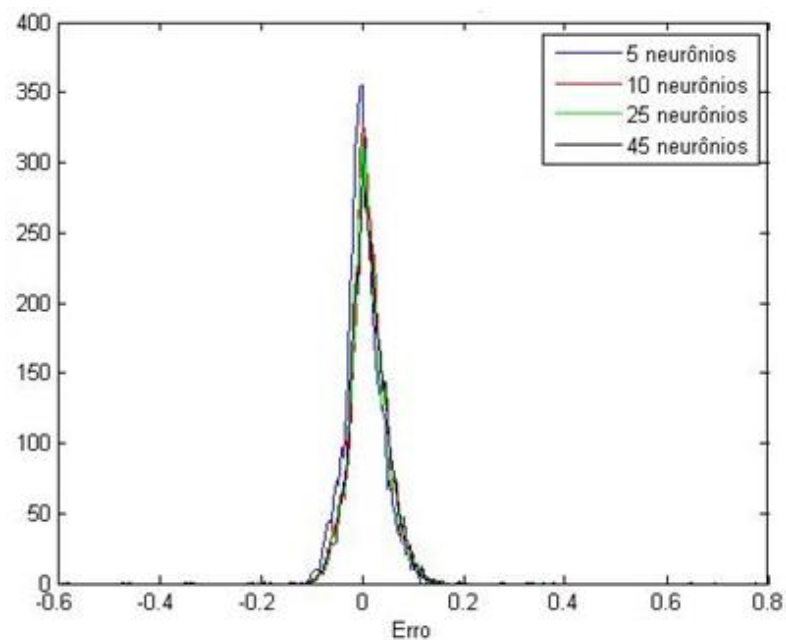


Figura 0.4 - fdp do erro dos modelos previsores para treinamento com $\sigma = 0.01$

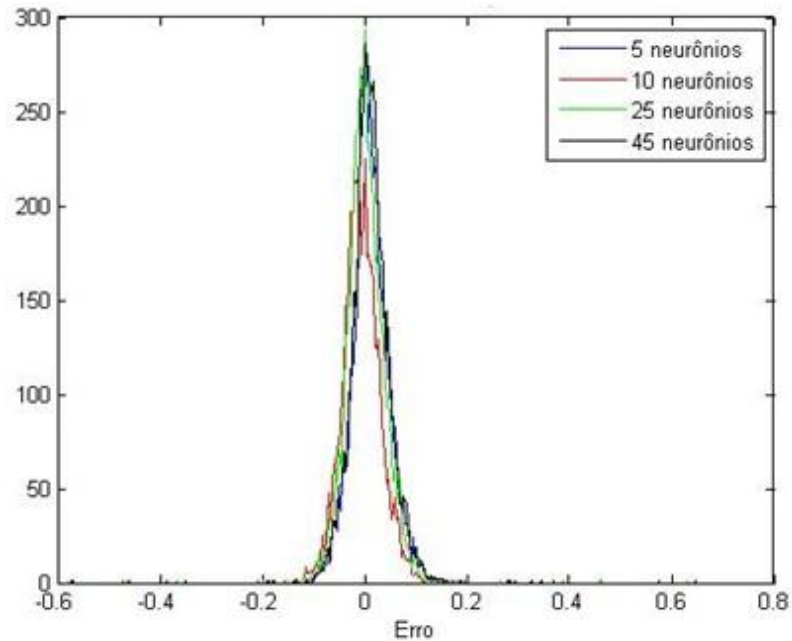


Figura 0.5 - fdp do erro dos modelos previsores para treinamento com $\sigma = 0.1$

Pela análise dos resultados pode-se verificar que os modelos que foram treinados com variância das funções gaussianas das janelas de Parzen igual a 0.01 apresentaram melhores resultados, no que diz respeito ao valor entropia e fdp do erro, sendo que entre estes modelos destaca-se a RNA com 5 neurônios na camada escondida que obteve uma fdp mais estreita, com maior número de erro próximos de zero. A importância da escolha do valor de sigma para o treinamento da RNA já foi mencionada no Capítulo 2. A estimativa de uma fdp com Janelas de Parzen pode originar uma função de bom comportamento e variação suave ou com muitos ótimos locais, dependendo da dimensão da janela usada. As Figuras 4.6 e 4.7 apresentam os histogramas do erro para dados de treinamento e validação para o modelo previsor com 5 neurônios na camada escondida e $\sigma = 0.01$ (melhor modelo).

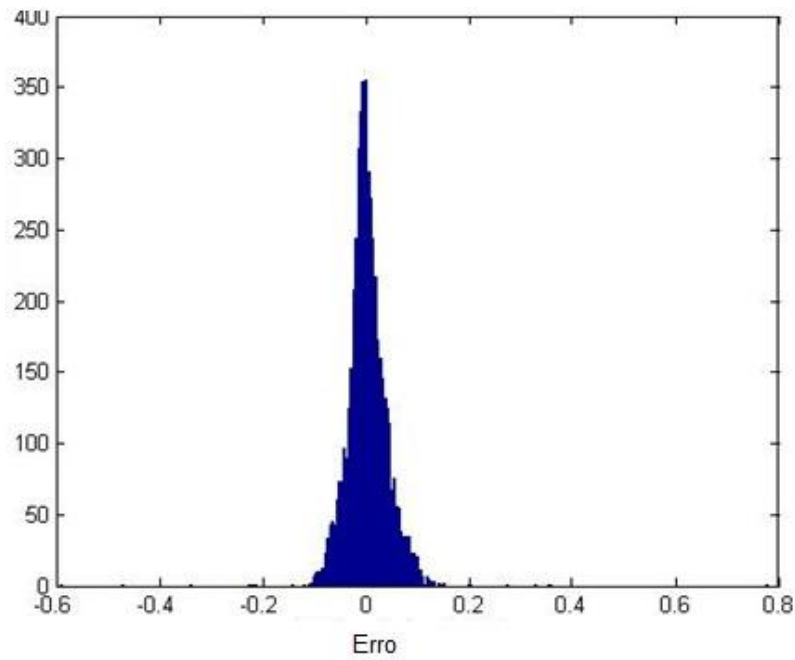


Figura 0.6 - Histograma do erro para dados de treinamento

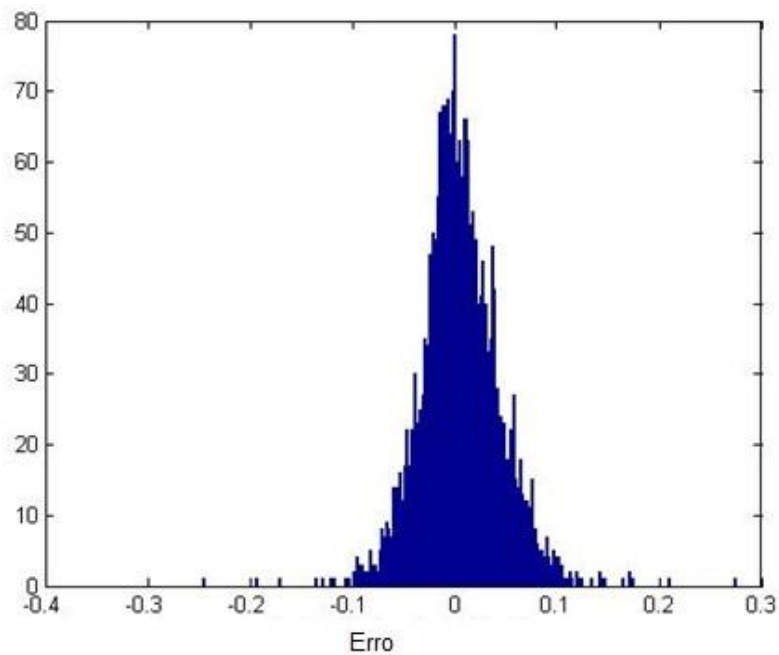


Figura 0.7 - Histograma do erro para dados de validação

As concessionárias de energia elétrica, para avaliar o desempenho da previsão gerada por um modelo utilizam como índice padrão o Erro Percentual Absoluto Médio (MAPE) . O índice MAPE é calculado por:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|valor\ previsto - valor\ real|}{valor\ real} 100\%$$

sendo N o número de padrões utilizados para desenvolvimento do modelo.

Para o modelo predictor desenvolvido o índice MAPE foi de 1.35%. Este índice será também utilizado posteriormente para comparação dos resultados entre os modelos propostos.

A Figura 4.8 apresenta os resultados do modelo predictor baseado na Entropia (MEE) para 24 horas de um dia em comparação com valor real da carga (alvo).

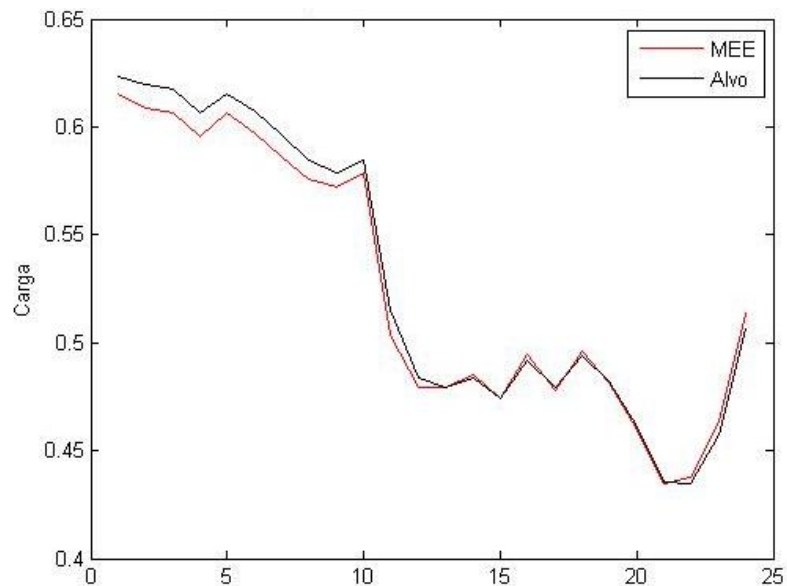


Figura 0.8 - Resultados do modelo predictor para 24 horas de um dia

4.5.2 MODELO PREVISOR BASEADO NO CRITÉRIO DA CORRENTROPIA

Para o desenvolvimento do modelo predictor baseado na MLP treinada pelo critério da Correntropia é importante destacar que diversas topologias (de acordo com a Figura 4.2) foram treinadas e testadas exaustivamente variando-se o número de neurônios da camada escondida e o valor da abertura da janela de Parzen. A Tabela 4.3 apresenta alguns dos resultados desses diversos treinamentos. Para cada variação do número de neurônios na camada oculta (5, 10, 25 e 45), foram testados valores distintos da variância/sigma (0.3, 0.01 e 0.1) das funções gaussianas das janelas de Parzen. Como critério de parada para os treinamentos adotou-se a variação do valor da correntropia ao longo do treinamento: quando este valor já não variava ao longo das iterações, o treinamento era finalizado.

Tabela 0.3 - Resultado dos treinos da RNA baseada no critério da Correntropia

Correntropia Sigma 0.3	5 neurônios na camada oculta	10 neurônios na camada oculta	25 neurônios na camada oculta	45 neurônios camada oculta
Nº de iterações	4000	4500	6000	2000
Valor da correntropia	-0.102	-0.109	-0.11	-0.11
Erro Médio (treino)	0.0016	0.0018	0.0016	0.0016
Erro Médio (validação)	0.0015	0.0017	0.0015	0.0014
Correntropia Sigma 0.01	5 neurônios na camada oculta	10 neurônios na camada oculta	25 neurônios na camada oculta	45 neurônios camada oculta
Nº de iterações	1000	1000	500	500
Valor da correntropia	-0.9350	-0.7077	-0.9103	-0.9386
Erro Médio (treino)	0.0016	0.0016	0.0016	0.0016
Erro Médio (validação)	0.0015	0.0014	0.0014	0.0014
Correntropia Sigma 0.1	5 neurônios na camada oculta	10 neurônios na camada oculta	25 neurônios na camada oculta	45 neurônios camada oculta
Nº de iterações	500	500	500	500
Valor da correntropia	-0.49	-0.47	-0.46	-0.47
Erro Médio (treino)	0.0016	0.0016	0.0016	0.0017
Erro Médio (validação)	0.0014	0.0014	0.0014	0.0014

Os gráficos da função densidade de probabilidade (fdp) do erro para os modelos apresentados na Tabela 4.3 são apresentados nas Figuras 4.9, 4.10 e 4.11 para cada valor de sigma e variando-se o número de neurônios da camada oculta.

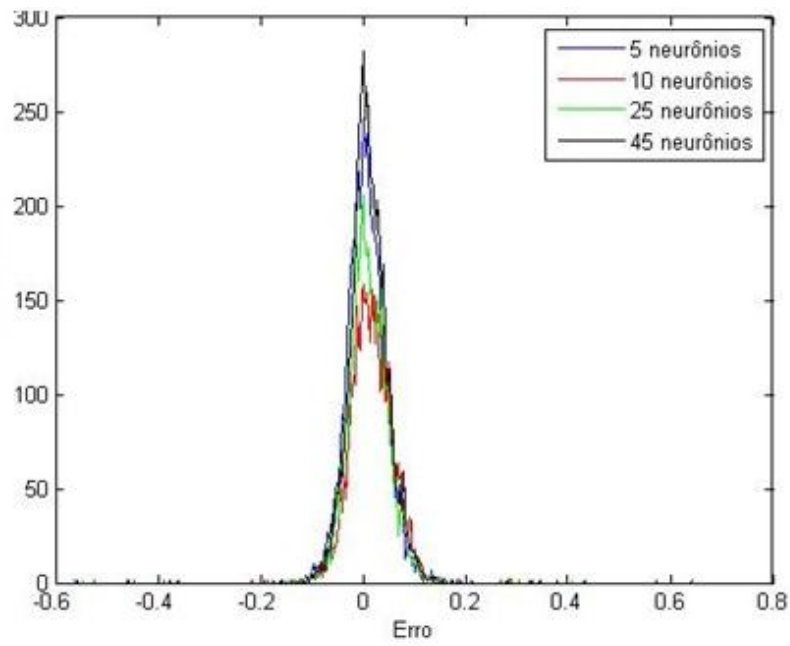


Figura 0.9 - fdp do erro dos modelos previsores para treinamento com $\sigma = 0.3$

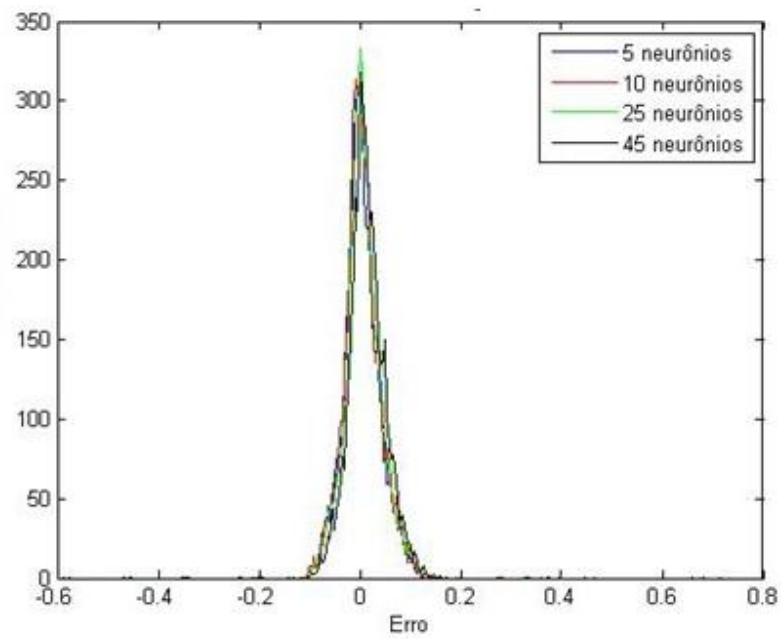


Figura 0.10 - fdp do erro dos modelos previsores para treinamento com $\sigma = 0.01$

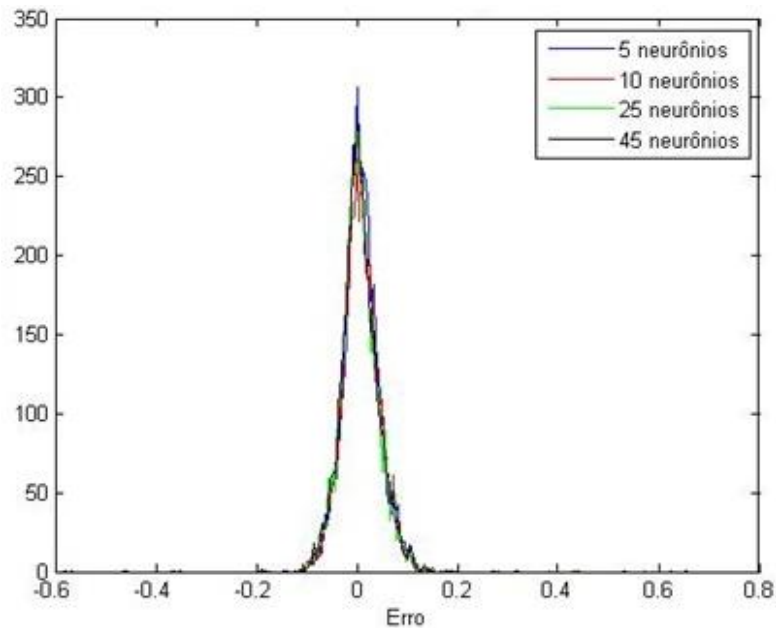


Figura 0.11 - fdp do erro dos modelos prevores para treinamento com $\sigma = 0.1$

Pela análise dos resultados pode-se verificar que os modelos que foram treinados com variância das funções gaussianas das janelas de Parzen igual a 0.01 apresentaram melhores resultados no que diz respeito ao valor Correntropia e fdp do erro, sendo que entre estes modelos destaca-se a RNA com 25 neurônios na camada escondida. A importância da escolha do valor de sigma para o treinamento da RNA já foi mencionada no Capítulo 2. As Figuras 4.12 e 4.13 apresentam os histogramas do erro para dados de treinamento e validação para o modelo previsor com 25 neurônios na camada escondida e $\sigma = 0.01$ (melhor modelo).

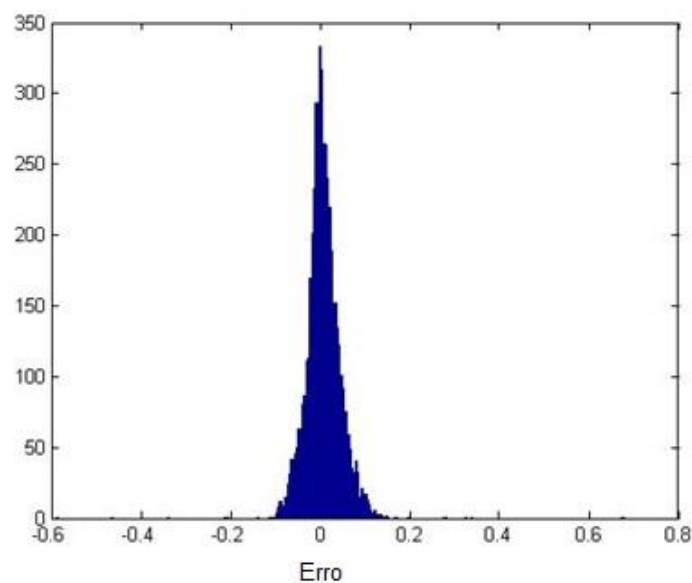


Figura 0.12 - Histograma do erro para dados de treino

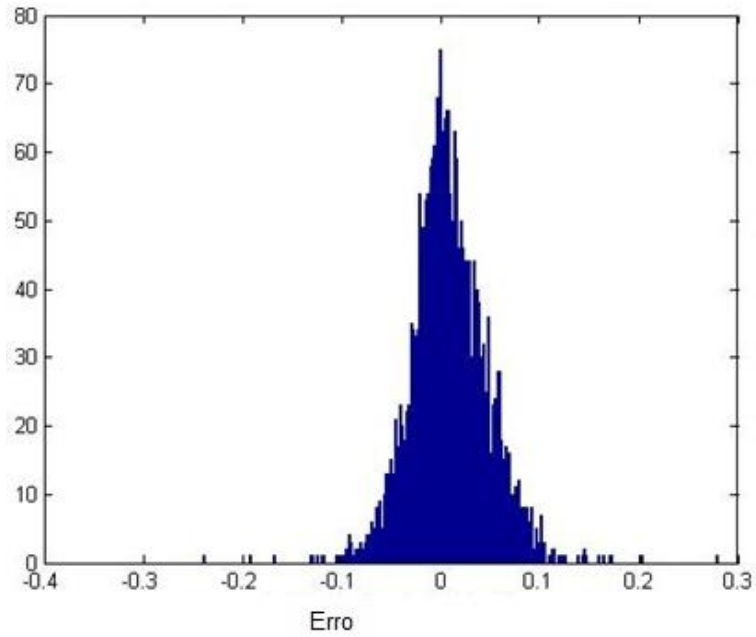


Figura 0.13 - Histograma do erro para dados de validação

Para o modelo predictor desenvolvido o índice MAPE foi de 1,39%. Este índice será também utilizado posteriormente para comparação dos resultados entre os modelos propostos.

A Figura 4.14 apresenta os resultados do modelo predictor baseado na correntropia (MEC) para 24 horas de um dia em comparação com valor real da carga (alvo).

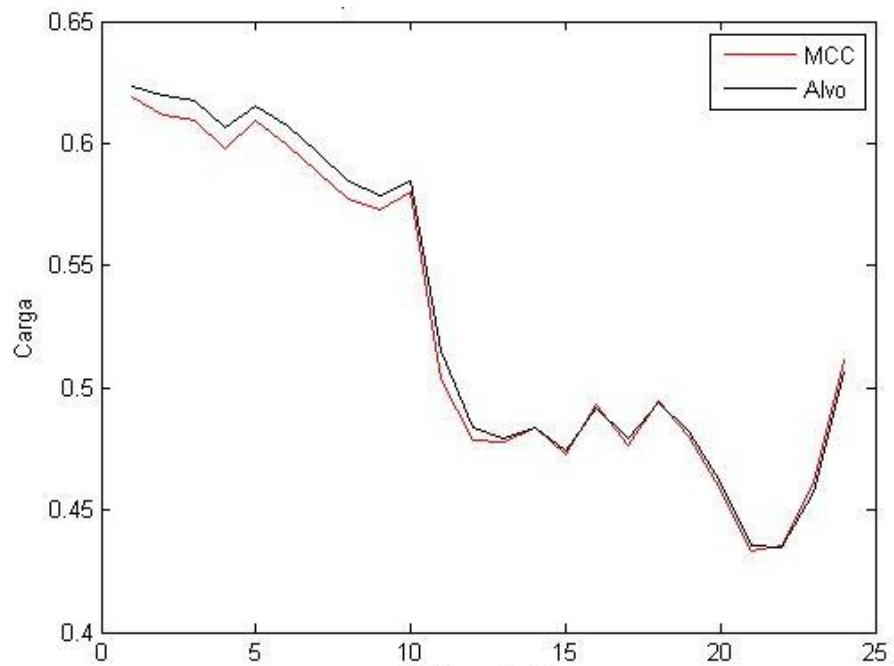


Figura 0.14 - Resultados do modelo predictor para 24 horas de um dia

4.6 MODELO PREVISOR BASEADO NO MSE

Para o desenvolvimento do modelo predictor baseado na MLP treinada pelo erro médio quadrático (MSE) diversas topologias (de acordo com a Figura 4.2) foram treinadas e testadas exaustivamente variando-se o número de neurônios da camada escondida. A Tabela 4.4 apresenta alguns dos resultados desses diversos treinamentos. Na tabela são apresentados além do valor do Erro Médio Quadrático para dados de treino e validação, o valor da entropia e índice MAPE para cada modelo.

Para o treino baseado no MSE foi utilizado o toolbox de Redes Neurais do MATLAB. Como critério de parada adotou-se a variação do MSE ao longo do treinamento: quando este valor já não variava ao longo das iterações, o treinamento era finalizado.

Tabela 0.4 - Resultados dos treinos da RNA baseada no critério MSE

MSE	5 neurônios na camada oculta	10 neurônios na camada oculta	25 neurônios na camada oculta	45 neurônios na camada oculta
Nº de iterações	12	11	14	10
Valor da entropia	-0.1172	-0.1170	-0.1170	-0.1171
Erro de treino	0.0014	0.0016	0.0015	0.0016
Erro de validação	0.0015	0.0015	0.0015	0.0015
MAPE	5,5%	5.7%	5,63%	5,52%

Pela análise dos resultados pode-se verificar que entre os modelos destaca-se a RNA com 5 neurônios na camada escondida. As Figuras 4.15 e 4.16 apresentam os histogramas do erro para dados de treinamento e validação para o modelo predictor baseado no critério clássico do MSE com 5 neurônios na camada escondida.

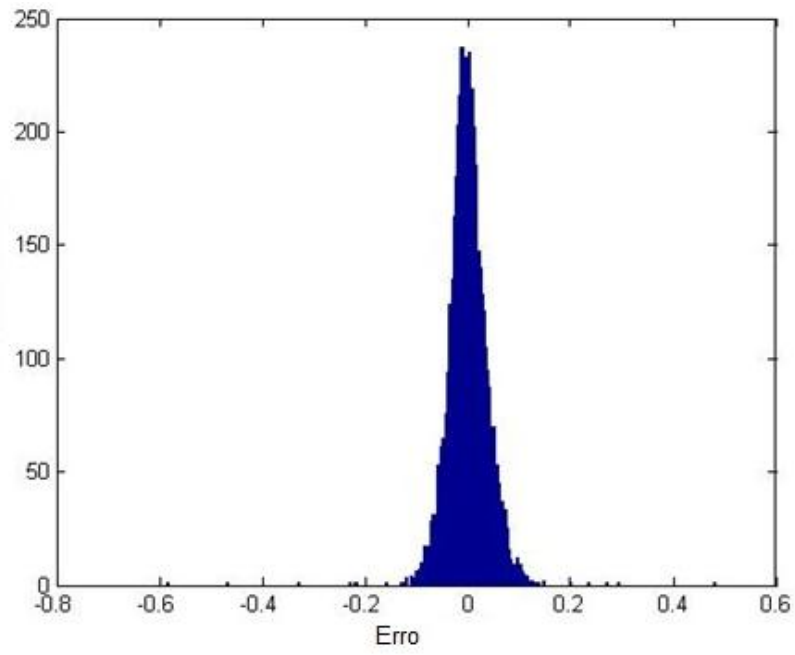


Figura 0.15 - Histograma do erro para dados de treino

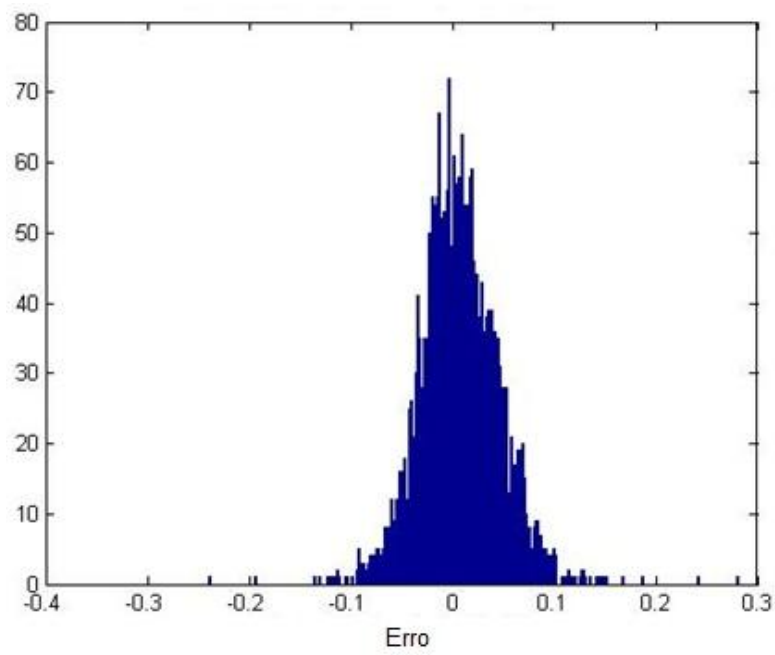


Figura 0.16 - Histograma do erro para dados de validação

4.7 ANÁLISE COMPARATIVA DOS MODELOS PREVISORES PROPOSTOS

A Tabela 4.5 apresenta um resumo dos resultados obtidos para os 3 modelos.

Tabela 0.5 - Resultados dos modelos previsores desenvolvidos

	MEE (Entropia)	MEC (Correntropia)	MSE (Erro médio quadrático)
Topologia	2-5-1	2-25-1	2-5-1
Nº de iterações	1000	500	12
Valor (entropia ou correntropia)	-0.93	-0.91	-0.1172 (valor da entropia)
Erro de treino (MSE)	0.0016	0.0016	0.0015
Erro de validação (MSE)	0.0014	0.0014	0.0015
MAPE de treino	1,35%	1,39%	5,5%
MAPE de validação	1,36%	1,40%	5,18%

Pode ser verificado que em relação à medida do Erro médio quadrático (MSE), os três modelos apresentaram valores próximos. O MSE é uma medida de dispersão do erro e muito sensível a valores de erros elevados. Em uma análise dos valores máximos e mínimos dos erros, verificou-se que para os 3 modelos os valores são próximos, justificando assim a proximidade do erro médio quadrático para os 3 modelos.

Em relação ao índice MAPE pode ser observado que os dois modelos treinados pela Teoria da Informação apresentaram resultados superiores ao modelo treinado pelo MSE. O índice MAPE é um índice mais adequado para avaliação de modelos previsores pois oferece uma medida do comportamento médio da variável erro.

Considerando os resultados dos modelos previsores já apresentados na literatura, os 3 modelos propostos apresentaram resultados do índice MAPE que podem ser considerados adequados para o problema de previsão de cargas a curto prazo. Considerando todos os trabalhos analisados, a média do índice MAPE é de 4,75% (valor mínimo 1,5% e valor máximo 8%). Como pode ser observado, os modelos MEE e MCC apresentaram valores de

índice MAPE abaixo dos valores dos modelos de previsão já propostos enquanto que o MSE ficou com valor um pouco acima da média.

Para análise da distribuição dos erros dos modelos propostos, as Figuras 4.17 e 4.18 apresentam a fdp do erro obtidas para os três modelos, para os dados de treino e validação respectivamente.

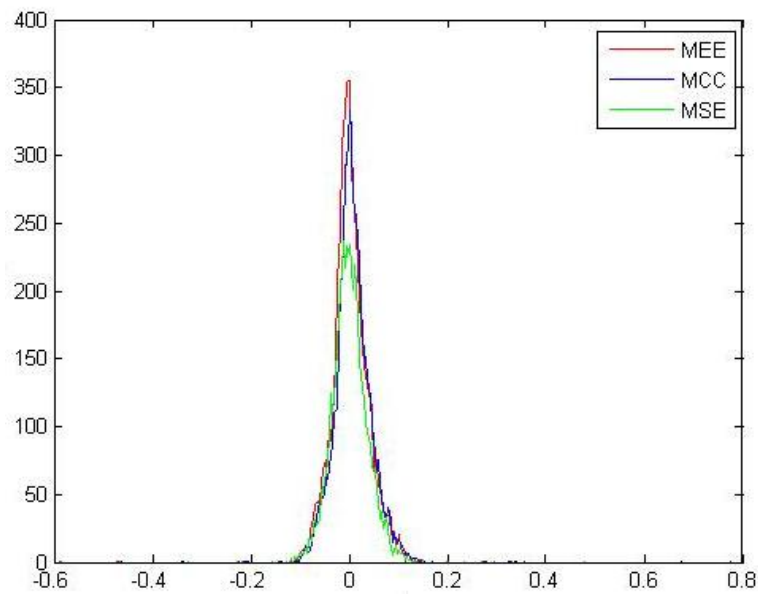


Figura 0.17 - Comparação da fdp entre os modelos para dados de treinamento

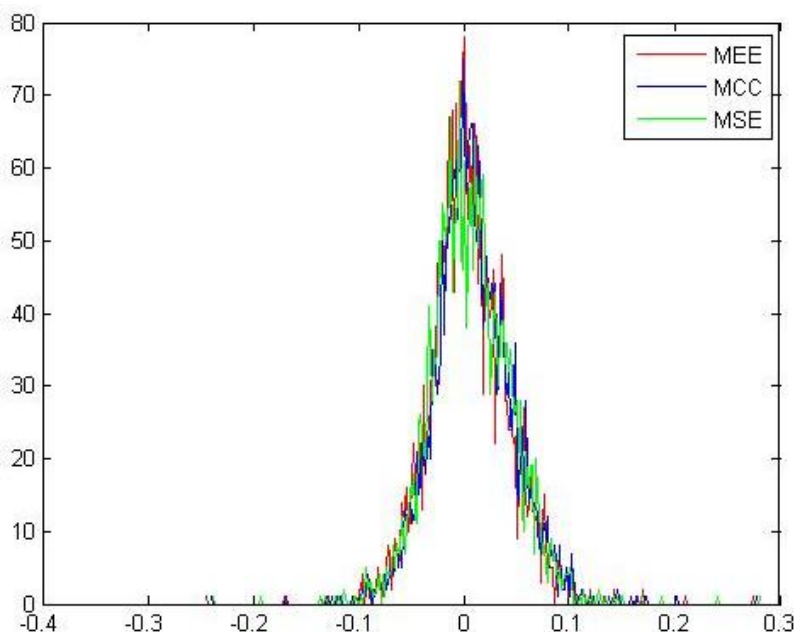


Figura 0.18 - Comparação da fdp entre os modelos para dados de validação

Pela análise das distribuições dos erros dos três modelos, pode ser verificado que se obteve uma fdp mais estreita para os modelos MEE e MCC, o que sinaliza um maior número

de erros próximos de zero do que o modelo MSE. Isto já era o esperado, pois com o treino pelo critério da entropia ou correntropia tem-se o acúmulo máximo de informação na estrutura dos pesos da rede (parâmetros ajustáveis durante treinamento) e apenas informação residual na distribuição do erro (distribuição do erro com maior frequência de valores próximos de zero – aproximando-se de uma Função de Dirac).

A Figura 4.19 apresenta para análise, o valor real da carga para 24 horas de um dia (alvo) e os valores previstos pelos três modelos.

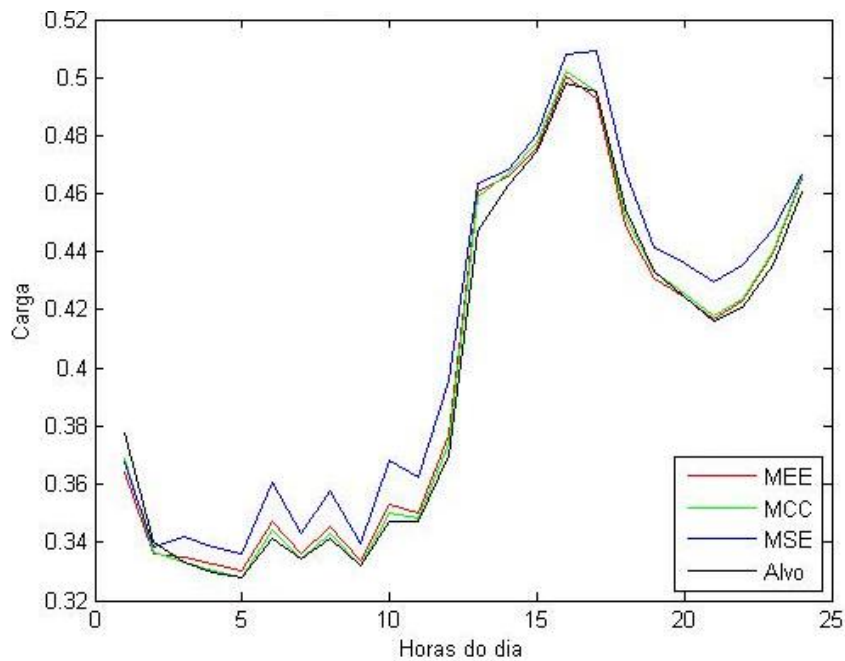


Figura 0.19 - Resultados para 24 horas do dia

Pode ser verificado que as curvas obtidas pelas redes treinadas pelo critério da entropia e correntropia se aproximam mais da curva real. Pode haver pontos onde o MSE se comporta melhor, com menor erro, mais é a curva dos modelos MEE e MCC que apresentam um maior número de pontos com valores mais próximos dos valores reais. Este comportamento foi observado ao longo de todos os dados de treino e validação.

Outro ponto que pode ser destacado em relação aos modelos MEE e MCC diz respeito a propriedade de generalização das Redes Neurais. Da Figura 4.18 é possível perceber que os modelos propostos, MEE e MCC, apresentam uma melhor generalização quando comparados com o modelo MSE. Isto mostra que os modelos propostos MEE e MCC apresentam uma melhor capacidade de resposta para dados não utilizados durante o processo de aprendizado.

Para finalizar, conclui-se que baseado nos resultados apresentados os modelos MEE e MCC se mostraram superiores ao modelo MSE, evidenciando desta maneira a aplicabilidade da Teoria da Informação para o aprendizado de sistemas de previsão de carga baseados em séries temporais. A principal vantagem observada em relação a estes modelos diz respeito a distribuição do erro que apresenta um maior número de valores próximos de zero, mostrando-se assim sistemas previsores mais precisos.

Capítulo 5

Conclusões

Atualmente, o estudo dos problemas de análise, planejamento e operação de sistemas de energia elétrica é de extrema importância para o funcionamento do sistema, sendo que obter o comportamento da carga elétrica é fundamental para a garantia do fornecimento de energia aos consumidores de forma econômica, segura e contínua.

A previsão de carga é uma das principais funções na operação de um sistema de energia elétrica. Estimar a demanda de energia futura de forma correta é prerrogativa obrigatória no gerenciamento da produção e distribuição dessa energia.

As metodologias para previsão de carga evoluíram muito nos últimos anos, entretanto as técnicas tradicionais, utilizadas pelas concessionárias de energia, geralmente são adequadas para serem aplicadas em condições normais de funcionamento do sistema, entretanto deixam a desejar em situações onde ocorrem mudanças inesperadas nos parâmetros do ambiente. Relacionamentos extremamente complicados entre variáveis que podem levar a operações matemáticas complexas na hora da previsão é outro fator negativo destas técnicas.

Para tentar suprir as desvantagens das metodologias tradicionais de previsão, sistemas de previsão baseados em técnicas de Inteligência Computacional vêm sendo desenvolvidos. As Redes Neurais Artificiais vêm sendo bastante utilizadas, principalmente devido a sua alta capacidade para modelagem de problemas não lineares.

Apesar de todos os modelos de previsão de carga com resultados satisfatórios já desenvolvidos, a busca por modelos mais precisos ainda é um ponto de interesse de muitas concessionárias de energia. Considerando os modelos baseados em redes neurais, muitos estudos ainda podem ser desenvolvidos com o intuito de se encontrar ótimas estruturas de redes ou abordagens de treinamento convenientes que, possivelmente, possam melhorar a precisão da previsão.

Considerando então a importância da obtenção de modelos cada vez mais precisos para o problema de previsão de carga, este trabalho apresentou os resultados do estudo investigativo a respeito da aplicação de redes neurais artificiais com treinamento baseado na Teoria da Informação para o problema de previsão de carga a curto prazo.

Os resultados alcançados mostraram a eficiência dos modelos baseados no treinamento pelo critério da Entropia e Correntropia propostos e contribui para evidenciar a

aplicabilidade da Teoria da Informação para o aprendizado de sistemas de previsão de carga baseados em séries temporais.

O estudo realizado mostra que o uso das redes neurais treinadas através de critérios baseados na Teoria da Informação podem ser mais adequados para o desenvolvimento de sistemas previsores do que as redes neurais treinadas pelo critério tradicional baseado no erro médio quadrático (MSE). Através do uso da entropia e correntropia pode-se obter sistemas de previsão de carga com distribuições de erro mais próximas da função de Dirac, o que é o desejado para estes tipos de sistemas visto que o que se deseja são sistemas de previsão cada vez mais precisos, pois estas previsões são de extrema importância para as diversas atividades de planejamento e operação dos sistemas de potência.

É importante salientar que as afirmações em relação à superioridade do uso da teoria da Informação em relação ao MSE é realizada considerando até agora modelos para previsão a curto prazo (1 hora adiante). Pretende-se ainda, como trabalho futuro, realizar um estudo mais abrangente onde serão considerados modelos de previsão, por exemplo, para daqui a 24 horas. Estudos também serão realizados para melhorias nos algoritmos para treinamentos baseados no critério da Entropia e Correntropia onde se pretende considerar janelas de Parzen com largura variável durante as iterações. Esta ideia surgiu da observação, durante a fase de treino dos modelos, da influência da largura da Janela para o sucesso do aprendizado.

Referências Bibliográficas

- Abu-El-Magd, M. A; Sinha, N. K; 1981.** *Two new algorithms for on-line modelling and forecasting of the load demand of multinode power system.* IEEE Transaction on power apparatus systems, pp. 592-600.
- Al-Kandari, A. M.; Soliman, S. A. and El-Hawary, M. E. 2004.** *Fuzzy short-term electric load forecasting.* Electric Power & Energy Systems, Vol. 26, No. 2, pp. 111-122.
- Almeida, C.; Fishwich, P. A. and Tang, Z. 1991.** *Time series forecasting using neural network vs. Box-Jenkins methodology.* Simulation Councils, Inc., pp. 303-310.
- Altiparmak, F; Dengiz, B. e Smith, A. E. 2009.** *A General Neural Network Model for Estimating Telecommunications Network Reliability.* IEEE Transactions on Reliability, pp. 2-9.
- Amjad, N. 2006.** *Day-ahead price forecasting of electricity markets by a new fuzzy neural network.* IEEE Transactions on Power Systems, pp. 887-896.
- Asbury, C.E; 1975.** *Weather load model for electric demand energy forecasting.* IEEE Transaction on Power Apparatus System, pp. 1111-1116.
- Baczynski, D. and Parol, M. 2004.** *Influence of artificial neural network structure on quality of short-term electric energy consumption forecast,”* IEEE Proceedings Generation Transmission Distribution, Vol. 151, No. 2, pp. 241-245.
- Barzamini, R., Menhaj, M., Khosravi, A. e Kamalvand, SH. 2005.** *Short Term Load Forecasting for Iran National Power System and Its Regions Using Multilayer Perceptron And Fuzzy Inference Systems,* Proceedings of International Joint Conference on Neural Networks.
- Bessa, R; Miranda, V e Gama, J. 2008.** *Improvement in Wind Power Forecasting Based on Information Entropy-Related Concepts.*
- Bunn, D. W. e Farmer, E. D, 1985.** *Comparative models for electrical load forecasting,* Wiley, Belfast.
- Caines, P. E; 1987.** *Linear stochastic systems.* John Wiley & Sons.
- Campo, R; Ruiz, P; 1987.** *Adaptive weather sensitive short-term load forecasting.* IEEE transactions on power systems, pp. 592-600.

- Castro, A. e Miranda V. 2002.** *Mapping neural networks into rule sets and making their hidden knowledge explicit – Application to spatial load forecasting.* PSCC02 – 14 Power Systems Computation Conf.
- Cargil, T. F. e Meyer, R. A; 1971.** *Estimating the demand for electricity by time of day.* Applied Economics, nr3, pp.233-246.
- Chen, C. S; Tzeng, Y. M; Hwang, J. C. 1996;** *The application of artificial neural networks to substation load forecasting.* Electric Power Systems Research Vol. 38, No. 2, pp. 153-160.
- Chogumaira, E.N.; Hiyama, T.; Elbaset, A.A. 2010.** *Short-term load forecasting using Dynamic Neural Networks.* Power and Energy Engineering Conference Asia-Pacific, pp. 1-4.
- Christiaanse, W. R; 1971.** *Short-term load forecasting using general exponential smoothing.* IEEE Transaction on Power Apparatus System, pp. 900-910.
- Csiszar, I; Korner, J. 1981.** *Information Theory: Coding Theorems for discrete memoryless system.* New York Ademic.
- Dehdashti, A. S; Tudor, J. R; Smith, M. C; 1982.** *Forecasting of hourly load by pattern recognition: a deterministic approach.* IEEE Transactions on Power Apparatus Systems, pp. 3290-3294.
- Drapper, N. R. e Smith, H, 1981.** *Applied Regression Analysis.* John Wiley & Sons Inc.
- Erdogmus D. e Principe J.C., 2002.** *Generalized Information Potential Criterion for Adaptive System Training.* IEEE Transactions on Neural Networks, vol. 13, no. 5, pp. 1035-1044.
- Gross, G. e Galiana, F. D, 1987.** *Short term load forecasting.* Proceedings of the IEEE, Vol. 75, No. 12, pp. 1558-1573.
- Gupta, P. C; Yamada, K; 1972.** *Adaptive short-term forecasting of hourly loads using weather information.* IEEE Transaction on Power Apparatus System, pp. 2085-2094.
- Gupta, P. C; 1971.** *A stochastic approach to peak power demand forecasting in electric utility systems.* IEEE Transaction on Power Apparatus System, pp. 824-831.
- Hagan, M. T; Klein, R; 1978.** *On-line maximum likelihood estimation for load forecasting.* IEEE Transaction on System, Man and Cybernetics, pp. 711-715.
- Harun, M.H.H.; Othman, M.M.; Musirin, I. 2010.** *Short-term load forecasting using artificial neural network based multiple lags and stationary time series.* Power Engineering and Optimization Conference, pp. 363-370.
- Haykin, Simon. 2001.** *Redes Neurais: Princípios e Prática.* 2ª Edição. Porto Alegre: Bookman, 2001.

- Heinemann, G. T; Nordman, D. A; Plant, E. C; 1966.** *The relationship between summer weather and summer loads – a regression analysis.* IEEE Transaction on Power System, pp. 1144-1154.
- Hippert, H. S; Pedreira, C. E; Souza, R. C, 2001.** *Neural networks for short-term load forecasting: Are view and evaluation.* IEEE Transactions on Power Systems, Vol. 16, No. 1, pp. 44-55.
- Hornik, K; Stinchcombe, M; White, H. 1989.** “*Multilayer feedforward networks are Universal Approximators*”, Neural Network, Vol. 2, pp 359-366.
- Huang S.J; Shih K.R; 2003.** *Short term load forecasting via ARMA Model identification including non-Gaussian process considerations.* IEEE Transactions on Power Systems, pp. 673
- Jabbour, K; Riveros, J.F; Landesbergen, D; Meyer, W; 1988.** *Alfa: Automated Load Forecasting Assistant.* IEEE Transaction on Power system, pp. 908-914.
- Jacobs, R. A. 1998.** *Increase Rates of Convergence through Learning Rate Adaptation.* Neural Networks. Vol. I, pp. 295-307.
- Krose, B. e Smagt, P. V. 1996.** *An introduction to neural networks.* Online Book, University of Amsterdam, Eighth Edition.
- Liao G.C; Tsao T.P; 2006.** *Application of a fuzzy neural network combined with a chaos genetic algorithm and simulated annealing to short-term load forecasting.* IEEE Transactions on Evolutionary Computation. pp. 330-340.
- Liao, G. C; Tsao, T.P; 2004.** *Application of fuzzy neural networks and artificial intelligence for load forecasting.* Electric Power Systems Research, Vol. 70, No. 3, pp. 237-244.
- Lijesen, D. P; Rosing, J; 1971.** *Adaptive forecasting of hourly loads based on load measurements and weather information.* IEEE Transaction on Power Apparatus System, pp. 1757-1767.
- Liu, K.; Subbarayan, S.; Shoults, R. R.; Manry, M. T.; Kwan, C.; Lewis, F. L. and Naccarino, J; 1996.** *Comparison of very short-term load forecasting techniques.* IEEE Transactions on Power Systems, Vol. 11, No. 2, pp. 877-882.
- Liu, W; Pokharel, P; Principe, J. 2007.** *Correntropy: properties and applications in non-Gaussian signal processing.* Accepted in IEEE Transactions on Signal Processing.
- Mendel, J.M and Maclaren, R.W. 1970.** *Reinforcement learning control and patern recognition systems,* in Adptive, learning an Pattern Recognition Systems: Theory and Applications, New York: Academic Press.
- Metaxiotis, K.; Kagiannas, A.; Askounis, D. and Psarras, J. 2003.** *Artificial intelligence in short term electric load forecasting: A state-of-the-art survey for the researcher.* Energy Conversion and Management, Vol. 44, No. 9, pp. 1525-1534.

- Moghram, I. e 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.
- Murto, P. 1998.** *Neural network models for short-term load forecasting.* Masters Thesis, Department of Engineering Physics and Mathematics, Helsinki University of Technology.
- Ning, Y; Liu, Y; Ji, Q. 2010.** *Bayesian BP-Neural Network based short-term load forecasting for power system.* International Conference on Advanced Computer Theory e Engineering (ICACTE), pp. v2-89 – v2-93.
- O'Donovan, T. M. 1983.** *Short Term Forecasting: An introduction to the Box-Jenkins approach,* John Wiley & Sons.
- Osman, Z.H. Awad, M.L. Mahmoud, T.K. 2009.** *Neural Network based approach for short-term load forecasting.* IEEE Power System Conference and Exposition PSCE'09, pp. 1.
- Park, D. C.; El-Sharkawi, M. A; Mark II, R. J. 1991.** *Electric load forecasting an artificial neural network,* IEEE Transactions on Power Systems, Vol. 6, No. 2, pp. 442-449.
- Parzen, E. 1962.** *On the estimation of a probability density function and the mode.* Annals Math. Statistics, v.33, 1962, p. 1065.
- Principe J. C. e Xu D. 1999.** *Introduction to information theoretic learning.* Proc. International Joint Conference on Neural Networks (IJCNN'99), Washington DC, USA, 10-16 July 1999, pp. 1783-1787.
- Rahman, S; Bahtnagar, R; 1988.** *An expert system based algorithm for short-term load forecasting.* IEEE Transactions on Power System, pp. 392-399.
- Renyi, A. 1976.** *Some Fundamental Questions of Information Theory.* Selected Papers of Alfred Renyi, Vol 2, pp. 526-552, Academia Kiado.
- Sfetsos, A. 2003.** *Short-term load forecasting with a hybrid clustering algorithm.* IEEE Proceedings Generation, Transmission, and Distribution, Vol. 150, No. 3, pp. 257-262.
- Shannon, C.E. 1948.** *A Mathematical Theory of Communications.* Bell Systems Technical Journal, vol. 27, pp. 379-423, pp. 623-656.
- Simpson, P. K. 1989.** *Artificial neural systems: Foundations, paradigms, applications, and implementations.* New York, Pergamon Press.
- Sun W; Lu J.C; He Y.J; Li J.Q; 2005;** *Application of neural network model combining information entropy and ant colony clustering theory for short-term load forecasting.* Proceedings of International Conference on Machine Learning and Cybernetics.
- Swarup, K. S. and Satish, B. 2002.** *Integrated ANN approach to forecast load.* IEEE Computer Applications in Power, Vol. 15, No. 1, pp. 46-51.

Thompson, R.P. 1976. *Wather sensitive electric demand and energy analysis on a large geographically diverse power system application to short-term hourly electric demand forecasting.* IEEE Transaction on Power System, pp. 385-393.

Toyoda, J; Chen, M. S; 1970. *An application of state estimation to short-term load forecasting.* IEEE – Transaction on Power Apparatus System, pp. 1678,1687.

Uppendar, J.; Gupta, C.P.; Singh, G.K.; Ramakrishna, G. 2010; *PSO and ANN-based fault classification for protective relaying.* Generation, Transmission & Distribution, IET. pp. 1197-1212.

Vemuri, S; Huang, W. L; Nelson, D. J; 1981. *On-line algorithms for forecasting hourly loads of an electric utility.* IEEE Transaction on Power Apparatus System, pp. 3775-3784.

Wang, L. X; 1997. *A course in fuzzy – systems and control.* Prentice-Hall International, Inc.

Wang, X.; Hatziargyrou, N; Tsoukalas, L. H. 2002. *A new methodology for nodal load forecasting in deregulated power systems.* IEEE Power Engineering Review, Vol. 22, No. 5, pp. 48-51, 2002.

Wasserman, P. D. 1989. *Neural computing - theory and practice.* Van Nostrand Reinhold.

Wei, L; Zheng-gang, Z. 2009. *Based on time sequence of ARIMA model in the application of short term electricity load forecasting.* International Conference on Research Challenges in Computer Science, pp. 11.

Weifeng, L; Pokharel, P; Principe, J. 2006. *Error Entropy, correntropy and m-estimation.* IEEE Signal Processing Society – Workshop on Machine Learning for Signal Processing. pp. 179-184.

Xinbo, Z e Jinsai, C. 2011. *Short-term Power System load forecasting based on improved BP artificial neural network.* IEEE International Conference on Computer Science and Automation Engineeering, pp. 14-17.

Yun, Z; Quan, Z; Caixin S. 2008. *RBF Neural Network and ANFIS-Based Short-Term Load Forecasting Approach in Real-Time Price Environment.* IEEE Transaction on Power Systems, pp. 853-858.

Apêndice A

Algoritmo de retropropagação baseado no critério MEE

Algorithm

INPUT: Vetor de entradas, Vetor saída desejada, tamanho da abertura da janela σ , numero de exemplos do vetor de entrada N , numero de camadas l

OUTPUT: Pesos W , bias B da rede neural, erro, entropia

begin

for $s \in [1:N]$ **do**

for $t \in [1:N]$ **do**

1. Avaliação dos gradientes locais para a camada de saída l

$$\delta^l(e(s)|e(t)) = -IF(e(s)|e(t)) \cdot f'(net^l(s))$$

$$\delta^l(e(t)|e(s)) = -IF(e(t)|e(s)) \cdot f'(net^l(t))$$

2. Avaliação dos gradientes locais para a camada de índice o de $l-1$ até 1 e para o neurônio de índice j

$$\delta_j^o(e(s)|e(t)) = f'(net_j^o(s)) \cdot \sum_{k=1}^{m_o+1} \delta_k^{o+1}(e(s)|e(t)) w_{kj}^{o+1}$$

$$\delta_j^o(e(t)|e(s)) = f'(net_j^o(t)) \cdot \sum_{k=1}^{m_o+1} \delta_k^{o+1}(e(t)|e(s)) w_{kj}^{o+1}$$

3. Para cada camada do índice o que vai desde 1 até l avaliar a atualização dos pesos.

$$\Delta w_{ji}^o = \eta \left(\delta_j^o(e(s)|e(t)) \cdot O_i^{o-1}(s) + \delta_j^o(e(t)|e(s)) \cdot O_i^{o-1}(t) \right)$$

end for

end for

end begin

end Algorithm

Apêndice B

Algoritmo de retropropagação baseado no critério MCC

Algorithm

INPUT: Vetor de entradas, Vetor saída desejada, tamanho da abertura da janela σ , numero de exemplos do vetor de entrada N , numero de camadas l , numero de épocas e .

OUTPUT: Pesos W , bias B da rede neural, erro, correntropia

begin

for $s \in [1:N]$ **do**

for $t \in [1:N]$ **do**

 1. Avaliação dos gradientes locais para a camada de saída l

$$\delta^l(e(s)) = K'_\sigma(y - d) \cdot f'(net^l(s))$$

 2. Avaliação dos gradientes locais para a camada de índice o de $l-1$ até 1 e para o neurônio de índice j

$$\delta_j^o(e(s)) = f'(net_j^o(s)) \sum_{k=1}^{m_o+1} \delta_k^{o+1}(e(s)) w_{kj}^{o+1}$$

 3. Para cada camada do índice o que vai desde 1 até l avaliar a atualização dos pesos.

$$\Delta w_{ji}^o = \eta (\delta_j^o(e(s)) \cdot O_i^{o-1}(s))$$

end for

end for

end begin

end Algorithm