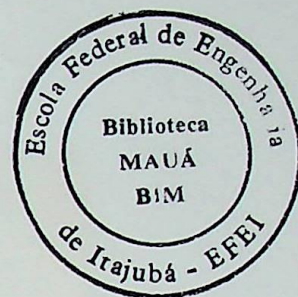


ESCOLA FEDERAL DE ENGENHARIA DE ITAJUBÁ

DISSERTAÇÃO DE MESTRADO



**PREVISÃO DA CARGA ELÉTRICA A CURTO-PRAZO
VIA REDES NEURAS**

CARLOS FERMÍN COLMÁN CRÚZ

ORIENTADOR : ALEXANDRE P. ALVES DA SILVA

CO-ORIENTADOR : GERMANO LAMBERT TORRES

**ITAJUBÁ - MG
1996**

CLASS. 621.316 : 519.711

CUTTER. C.955p

TOMBO. 841

EFEI / BIBLIOTECA

ESTE LIVRO DEVE SER DEVOLVIDO NA
ÚLTIMA DATA CARIMBADA .

19.8.96	17/11/2000
17/01/96	17/01/96
03/10/96	04/12/90
25/02/97	21.12.00
22.4.97	08.01.2001
24.6.97	25/01/01
14/07/97	13.2.01
14/07/97	14/07/97
31/07/99	15/02/2002
04.12.97	23/06/02
23/01/98	22/04/03
09/02/98	29.4.2003
04.9.98	09.5.2003
19/09/98	23.5.2003
25.9.98	27/5/05
20/10/98	
02/02/99	
20/02/99	
09.03.99	
25.3.99	
12.4.99	
31.10.2000	

AGRADECIMENTOS

Quero agradecer aos meus professores orientadores pelo apoio e pela força em todo momento, aos meus colegas do Grupo de Engenharia de Sistemas (GESis), à Escola Federal de Engenharia de Itajubá pela oportunidade oferecida, à CAPES pelo apoio econômico, aos meus pais pela compreensão e a todos meus amigos.

RESUMO

A previsão da carga é uma ferramenta útil para o operador do sistema na operação econômica e segura de um Sistema Elétrico de Potência.

O problema da previsão da carga elétrica a curto-prazo vem sendo estudado intensamente há muito tempo. Várias técnicas têm sido propostas nas décadas passadas .

Este trabalho tem por objetivo apresentar um estudo comparativo de previsão de carga a curto-prazo utilizando as técnicas de Redes Neurais Artificiais e Métodos Estatísticos Clássicos (Box & Jenkins). É verificada a influência de certos fatores na precisão da previsão, como: agrupamento, tendência, sazonalidade e a transformação Box-Cox.

Em Redes Neurais, é feita uma análise comparativa entre os algoritmos de Retropropagação de erro (Backpropagation) e GMDH (Group Method of Data Handling).

Para tal estudo, é utilizada a base de dados de cargas reais coletadas pela Hydro-Quebec do Canadá, contendo informações relacionadas a um barramento do sistema com carga predominantemente residencial. Estes valores foram medidos em uma base horária.

O trabalho também apresenta uma revisão das técnicas de previsão existentes citando as principais referências bibliográficas.

ABSTRACT

The load forecasting is an useful tool to help system operators in the economic and secure operation of electric power systems.

The short-term load forecasting problem have been studied for a long time. Several techniques have been lately.

The objective of this work is to present comparative study of short-term load forecasting using Artificial Neural Networks and classic statistics methods. The influence of several factors in the precision of the load forecasting is evaluated, such as data clustering, trending, sazonality and the Box-Cox transformation.

A comparative study between two Neural Networks training algorithms, Backpropagation and Group Method Data Handling, is presented.

A data base of real load collected by the Hydro-Quebec of Canada is used, including information of system busbar load with residential profile. The load values were read in a hour base.

This work also, presents a review of forecasting techniques with the corresponding bibliography.

ÍNDICE

	Pags.
Agradecimentos	i
Resumo	ii
Abstract	iii
Índice	iv
 Capítulo 1 - O estado da arte 	
1.1 - Introdução	1
1.2 - Definição e finalidade	2
1.3 - A importância da previsão a curto-prazo	2
1.4 - A previsão da carga elétrica a curto-prazo dentro do sistema gerenciador de energia	4
1.5 - A carga do sistema	5
1.5.1 - Fatores temporais	6
1.5.2 - Fatores meteorológicos	6
1.5.3 - Fatores aleatórios	7
1.6 - Classificação do tipo do modelo de carga utilizado	7
1.6.1 - Modelos de picos de carga	7
1.6.2 - Modelos de curva de carga	8
1.7 - Redes Neurais em previsão de carga	10
1.8 - Metodologia proposta	10
 Capítulo 2 - Métodos de previsão da carga elétrica 	
2.1 - Introdução	11
2.2 - Métodos quantitativos (estatísticos)	11
2.2.1 - Método da Regressão Linear Múltipla (RLM)	11
2.2.2 - Método de Alisamento Exponencial (MAE)	12
2.2.3 - Séries temporais estocásticas	12
2.2.3.1 - Conceitos básicos das séries temporais	12
2.2.3.2 - Método das séries temporais estocásticas	14
2.2.4 - Método de espaço de estado e filtro de Kalman (EEFK)	19
2.3 - Métodos qualitativos (baseados no conhecimento)	20
2.3.1 - Método baseado em sistemas especialistas	20
2.3.2 - Método baseado em Redes Neurais Artificiais (RNAs)	21
2.4 - Novas tendências	22



2.5 - Considerações práticas	22
2.5.1 - Formulação e seleção do modelo	23
2.5.2 - Filtragem dos dados	24
2.5.3 - Avaliação de desempenho	25

Capítulo 3 - Redes Neurais Artificiais

3.1 - Introdução	27
3.2 - Fundamento das redes neurais artificiais	28
3.3 - Histórico do desenvolvimento das redes neurais	30
3.4 - Características do desempenho das redes neurais	33
3.5 - Modelos de redes neurais	34
3.5.1 - Modelo multicamadas (feed-forward network)	35
3.5.2 - Modelo Hopfield	36
3.5.3 - Modelo de Kohonen	36
3.5.4 - Modelo Adaline e Madaline	37
3.5.5 - Modelo Link-Net	38
3.6 - Principais aplicações das redes neurais	40
3.7 - Comparação com outras técnicas de modelamento	40
3.8 - Dimensionamento da rede neural	42
3.9 - Condicionamento dos dados e influência do tamanho do conjunto de treinamento	43
3.10 - Problemas mais comuns envolvidas durante o processo de treinamento	44
3.11 - Treinamento da rede neural (Backpropagation)	45
3.12 - Algoritmo Backpropagation	46
3.13 - Algoritmo do GMDH "Group Method of Data Handling"	49

Capítulo 4 - Redes neurais em previsão de carga

4.1 - Introdução	56
4.2 - Revisão Bibliográfica	56
4.3 - Modelo proposto	58
4.3.1 - Influência do agrupamento	60
4.3.2 - Influência do destendenciamento	60
4.3.3 - Influência da sazonalização	61
4.3.4 - Influência da transformação Box-Cox	61
4.3.5 - Modelo predictor de redes neurais (feed-forward)	61
4.3.6 - Modelo linear de séries temporais	62
4.3.7 - Propriedades de um modelo linear	63
4.3.8 - Ajuste do modelo linear para uma série temporal	65
4.3.9 - Seleção do modelo	66
4.3.10 - Falha dos modelos lineares	66
4.3.11 - Modelo conexionista	67

Capítulo 5 - Testes e resultados

5.1 - Introdução	68
5.2 - Teste 1: Comparação entre o algoritmo de retropropagação do erro e GMDH sem agrupamento de dados	68
5.3 - Teste 2: Comparação da previsão com dados globais destendenciados e destendenciados+desazonalisados, utilizando o GMDH	70
5.4 - Teste 3: Comparação da previsão com dados globais destendenciados+desazonalisados e aplicando a transformação Box-Cox, usando o GMDH	71
5.5 - Teste 4: Comparação da previsão com dados destendenciados, destendenciados+desazonalisados aplicando a transformação Box-Cox, utilizando agrupamento de modelos de carga similares com o GMDH	72
5.5.1 - Caso 1: Agrupamento para dias sábados	72
5.5.2 - Caso 2: Agrupamento para dias domingos	75
5.5.3 - Caso 3: Agrupamento dias da semana	77
5.6 - Teste 5: Comparação dos resultados da previsão, considerando valores de cargas globais, destendenciados+desazonalisados e utilizando a transformação Box-Cox, com o modelo ARIMA	79
5.7 - Teste 6: Comparação dos resultados da previsão por agrupamento considerando dados destendenciados+desazonalisados e a transformação Box-Cox, utilizando o modelo ARIMA	80
5.7.1 - Caso 1: Agrupamento dias sábados	81
5.7.2 - Caso 2: Agrupamento dias domingos	82
5.7.3 - Caso 3: Agrupamento dias da semana	83
5.8 - Conclusão dos teste realizados	84
Tabelas comparativas	86

Capítulo 6

Conclusões	95
Propostas de futuros trabalhos	96
REFERÊNCIAS	97

CAPÍTULO 1

O ESTADO DA ARTE

1.1 Introdução

A previsão da carga elétrica constitui uma parte essencial no planejamento e na operação de um sistema elétrico de potência SEP. Um sistema de previsão adequado poderia reduzir de forma significativa os custos operacionais de um SEP. Nesse contexto foi introduzida a necessidade de melhorar os métodos de previsão. Várias técnicas e métodos têm sido investigados para abordar o problema nas últimas décadas.

O objetivo principal é prever a carga horária, diária, ou semanal que são necessárias para o planejamento da operação de um SEP [1]. Para efeito de estudo, a previsão da carga pode ser dividida em três (3) categorias principais: a curtíssimo prazo, a curto-prazo e a longo prazo.

a. Previsão a curtíssimo prazo

Nesta categoria o horizonte de previsão é de uns poucos minutos a uma hora à frente. Isto é necessário para controle e operação on-line de um sistema de potência, incluindo: despacho econômico e avaliação da segurança do sistema. Como exemplo da utilização desta previsão, pode-se citar a autorização para pequenos rearranjos do programa de manutenção preditiva em linhas, transformadores e unidades geradoras, bem como chaveamento de emergência no sistema.

b. Previsão a curto prazo

O horizonte da previsão a curto-prazo é de uma hora até uma semana à frente. Isto é necessário para a programação ótima das unidades geradoras (programação das partidas/paradas das unidades hidro-térmicas), controle das reservas operativas e a programação de intercâmbios com outras áreas do sistema interligado.

c. Previsão a longo prazo

Nesta categoria inclui-se previsão de picos de carga anual até 10 anos. Esta informação é utilizada no planejamento e dimensionamento de novos equipamentos e unidades geradoras.

1.2 Definição e Finalidade

Neste trabalho será abordada a previsão de carga a curto-prazo (PCCP). Doravante será utilizado o termo "curto-prazo" indicando previsão na ordem de horas. Os horizontes de previsão são da próxima hora até 168 horas (uma semana). O período de interesse na previsão de carga a curto-prazo é tipicamente de uma hora. Em adição à previsão horária dos valores das cargas do sistema, este trabalho inclui também previsão de:

- Picos máximos e mínimos instantâneos da carga do sistema
- Valores das cargas do sistema numa certa hora do dia
- Carga por barramento
- Carga por malha regional

1.3 A importância da previsão de carga a curto-prazo

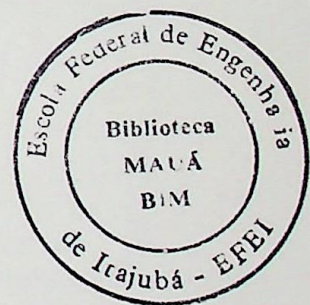
A PCCP é de vital importância para uma formulação econômica e segura da estratégia de operação do sistema

O principal objetivo é prover a carga prevista para:

- Programar a geração das unidades.
- Informar hora a hora seu valor ao operador do sistema.
- Avaliar a segurança do sistema de potência em qualquer instante.

Uma primeira aplicação seria determinar a programação mais econômica das unidades geradoras. Para sistemas puramente hidráulicos, a previsão da carga é requerida afim de determinar o volume de água vertido pelos reservatórios e os níveis de geração nas usinas.

Para sistemas puramente térmicos, a previsão da carga é necessária para a programação das unidades geradoras, com a finalidade de determinar o custo mínimo horário das partidas e paradas das unidades de modo a atender a carga prevista.



Para sistemas mistos (hidráulicos e térmicos), as cargas previstas são utilizadas pela função de coordenação hidrotérmica na programação da operação horária das várias unidades geradoras, de forma a minimizar os custos de produção.

A programação hidráulica, a programação das unidades geradoras térmicas, e a coordenação hidrotérmica são funções que requerem o conhecimento da carga prevista do sistema para o próximo dia ou próxima semana, a fim de determinar o menor custo de operação possível.

Uma segunda aplicação da PCCP é a avaliação preditiva da segurança do sistema de potência. A carga prevista do sistema é um dado essencial para a análise off-line da rede, de forma a detectar condições futuras sob as quais o sistema de potência pode estar vulnerável. Esta informação permite aos operadores tomar as ações corretivas necessárias (por exemplo: compra de energia, operações de chaveamento, rejeição de carga) para operar o sistema de forma segura e confiável.

A terceira aplicação da PCCP é fornecer ao operador do sistema informações sobre a mais recente carga prevista, levando-se em consideração as previsões climáticas e o comportamento aleatório da carga. Os operadores necessitam destas informações para operar o sistema de forma econômica e confiável. A figura 1.1 resume as principais aplicações da previsão da carga a curto-prazo.

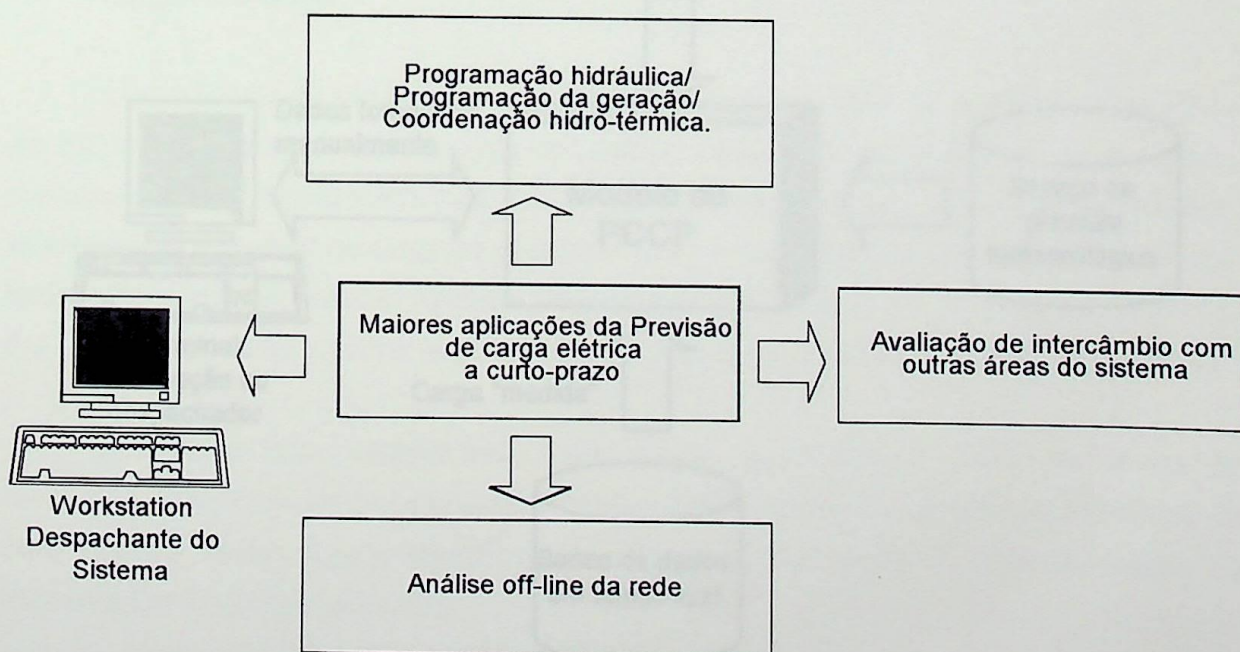


Fig. 1.1 - A PCCP é fornecida ao despachante do sistema como entrada para as funções de programação.

1.4 A previsão da carga elétrica a curto-prazo dentro do sistema gerenciador de energia

O procedimento empírico de previsão realizado pelo operador do sistema tem sido substituído por programas computacionais de previsão a curto-prazo dentro do sistema gerenciador de energia (SGE). Os componentes principais de um sistema de uma PCCP são: os modelos de previsão, banco de dados, e a interface homem-máquina.

Os modelos de previsão incluem a representação da carga do sistema e os algoritmos de previsão. As fontes de dados são os históricos de carga e um banco de dados climáticos com informações de: temperatura, velocidade do vento e humidade. A interface homem-máquina permite que o operador forneça e receba dados do modelo. Além disso, o modelo de PCCP é alimentado por dados medidos em tempo-real no sistema supervisionado e por um serviço de previsão meteorológica. A figura 1.2 ilustra os dados de entrada para o modelo de PCCP.

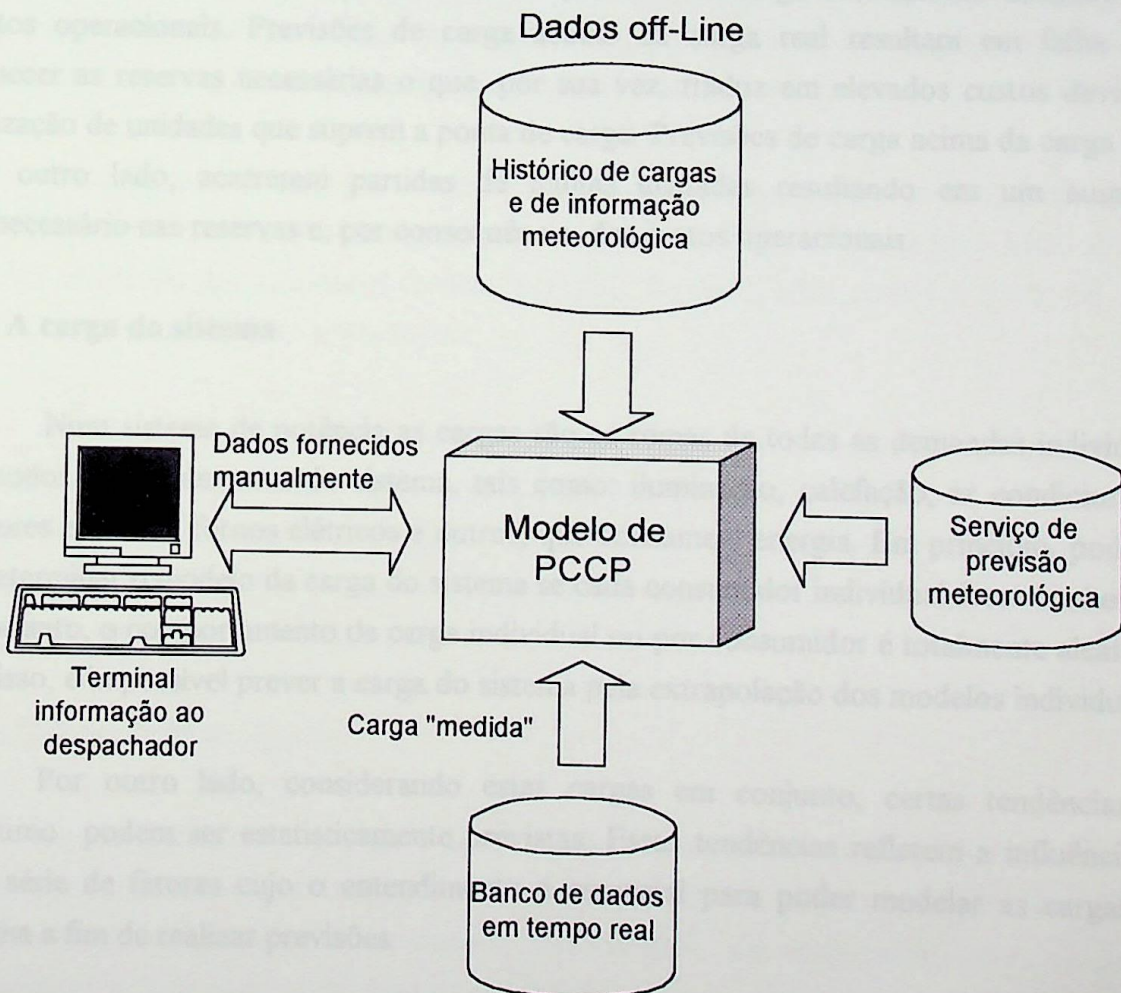


Fig. 1.2 - Entrada de fonte de dados para o modelo de previsão de carga

As saídas da previsão são fornecidas aos operadores do sistema e para as outras funções do sistema gerenciador de energia que necessitam desses valores de carga. A precisão da PCCP tem efeitos significativos na operação dos sistemas elétricos de potência e nos custos de produção. Os operadores do sistema devem antecipar o valor da carga do sistema para ter suficiente geração para satisfazer a demanda.

Ao mesmo tempo, níveis suficientes de reserva operativa e auxiliar são necessários para aliviar certos desvios inerentes à previsão. O custo das reservas é alto desde que as unidades que constituem as reservas não se encontrem plenamente carregadas, ou seja, estejam sub-utilizadas. As capacidades das reservas operativas e auxiliares são em certos níveis ditados pelas medidas necessárias de segurança e de confiabilidade desejadas para a operação do sistema de potência.

Assim pela redução do erro previsto, níveis de reserva podem ser reduzidos sem afetar a confiabilidade e segurança do sistema. Neste sentido, os custos de operação são diminuídos. Além disso, erros previstos na previsão da carga resultam no aumento dos custos operacionais. Previsões de carga abaixo da carga real resultam em falha para fornecer as reservas necessárias o que, por sua vez, traduz em elevados custos devido à utilização de unidades que suprem a ponta de carga. Previsões de carga acima da carga real, por outro lado, acarretam partidas de muitas unidades resultando em um aumento desnecessário nas reservas e, por consequência, dos custos operacionais.

1.5 A carga do sistema

Num sistema de potência as cargas são as somas de todas as demandas individuais em todos os barramentos do sistema, tais como: iluminação, calefação, ar condicionado, motores elétricos, fornos elétricos e outros, que consomem energia. Em princípio, poderia-se determinar o modelo da carga do sistema se cada consumidor individual fosse conhecido. Entretanto, o comportamento da carga individual ou por consumidor é totalmente aleatória. Por isso, é impossível prever a carga do sistema pela extrapolação dos modelos individuais.

Por outro lado, considerando estas cargas em conjunto, certas tendências de consumo podem ser estatisticamente previstas. Essas tendências refletem a influência de uma série de fatores cujo o entendimento é essencial para poder modelar as cargas do sistema a fim de realizar previsões.

O comportamento da carga a curto prazo é influenciado por certos fatores que podem ser classificados em três categorias principais [2]: fatores temporais, fatores climáticos e efeitos aleatórios. Para a previsão de carga a longo prazo, deve-se mencionar os fatores de natureza macro-econômica. A seguir discutiremos brevemente os efeitos de cada classe relacionada ao curto prazo.

1.5.1 *Fatores temporais*

Três principais fatores de origem temporal influenciam no modelo da carga, são eles: variações sazonais, periodicidade diária e /ou semanal e ocorrência de feriados.

As periodicidades diárias e /ou semanais da carga são manifestações das atividades sociais (ciclo trabalho - descanso) durante semanas típicas de determinada estação do ano. Assim existem modelos de carga bem definidos para períodos típicos da semana.

As variações sazonais refletem a influência das mudanças de estação ao longo do ano. Certas alterações no modelo da carga ocorre gradualmente em resposta às variações sazonais como: mudanças na temperatura e período de luz natural.

De qualquer modo, são os eventos sazonais que produzem as mais importantes modificações estruturais no modelo do consumo elétrico. Esses eventos são: início do ano escolar, alteração das atividades durante períodos de férias (ex.: períodos de Natal, Ano Novo), adoção do horário de verão, etc.

A ocorrência de feriados reflete o comportamento não rotineiro dos consumidores e normalmente produz o efeito de reduzir os valores da carga a níveis bem inferiores aos verificados em dias comuns. Além disso, observa-se que tal mudança de comportamento se antecipa e se prolonga em relação à ocorrência do feriado, produzindo um "alongamento" do mesmo ("inércia" de feriado).

1.5.2 *Fatores meteorológicos*

As condições meteorológicas também são responsáveis pelas variações no modelo da carga. Isto é devido à utilização em grande escala de diferentes equipamentos em diferentes épocas do ano, especialmente nas estações de verão e inverno, como: ar condicionados, aquecedores de ambientes e irrigadores para agricultura.

Em muitos sistemas, a temperatura é a mais importante variável no consumo de energia elétrica. Por exemplo, uma grande variação de temperatura pode causar mudanças significativas na carga requerendo assim modificações na programação das unidades geradoras. Para um sistema com uma geografia e clima não-uniforme, várias áreas pode ser necessária a fim de considerar os efeitos das variações climáticas na carga do sistema. A humidade é um outro fator que exerce influência sobre a carga do sistema de uma maneira similar à temperatura, particularmente em áreas quentes e úmidas. As chuvas também influenciam fortemente a carga devido às mudanças na temperatura. Outros fatores que influenciam o comportamento da carga são: velocidade e direção do vento, nebulosidade e nível de luminosidade.

1.5.3 Fatores aleatórios

Podem ser agrupados dentro desta categoria uma variedade de eventos que causam variações no modelo da carga sem ter relação com as condições meteorológicas. Um sistema de potência está continuamente sujeito a distúrbios aleatórios que afetam a carga global do sistema, além de um grande número de pequenos distúrbios como a operação de fornos elétricos, chaveamentos, problemas de tensão, entre outros. Estes distúrbios têm um carácter imprevisível e podem ocorrer a qualquer instante no sistema.

Também certos eventos como: entrada/saída de grandes consumidores, programas especiais de televisão (jogo de futebol, novelas), mudanças climáticas bruscas e greve geral são conhecidos a priori, porém os efeitos sobre a carga são incertos.

1.6 **Classificação do tipo do modelo de carga utilizado**

A classificação considera dois modelos básicos:

- Modelos de picos de carga.
- Modelos de curva de carga.

1.6.1 Modelos de picos de carga

O modelo do pico de carga horário ou semanal é gerado, na maioria dos casos, como uma função das condições meteorológicas. A hora do dia não desempenha papel algum em tais modelos que são tipicamente da forma:

$$p = B + F(W)$$

ou **pico de carga = carga base + componente dependente dos efeitos meteorológicos**

onde a carga base \mathbf{B} é uma componente média independente das condições meteorológicas a qual é adicionada a componente dependente $\mathbf{F}(\mathbf{W})$. As variáveis meteorológicas (\mathbf{W}) podem incluir a temperatura na hora do pico de carga ou uma combinação de temperaturas previstas e históricas.

Humidade, intensidade luminosa, velocidade do vento e precipitação têm sido também consideradas em tais modelos. A função $F(\cdot)$ é computada empiricamente e pode ser linear ou não-linear. Exemplos de modelos de picos de cargas podem ser encontrados em [3,4].

As vantagens do modelo de pico de carga são a sua estrutura simples e a baixa exigência de dados para inicialização e atualização. Os coeficientes do modelo são estimados através de regressão linear ou não-linear. As desvantagem de tais modelos são que eles não fornecem tanto a hora de ocorrência do pico quanto qualquer informação acerca do formato da curva de carga.

1.6.2 Modelos de curva de carga

Tais modelos descrevem a carga como uma série temporal discreta ao longo do intervalo de previsão. O intervalo de tempo considerado na amostragem da carga é tipicamente uma hora ou meia-hora, enquanto que os valores utilizados são energia consumida neste intervalo em MWh.

Muitas técnicas para previsão de carga descrevem a curva de carga uma vez que também inclui o pico de carga. Entretanto, dado que é difícil prever o pico de carga com muita precisão, modelos que combinam a curva e o pico da carga são desejáveis [4].

Basicamente, existem dois tipos de modelos de curva de carga: modelos hora-do-dia e modelos dinâmicos. Combinações destes dois tipos são também possíveis.

O modelo hora-do-dia define a carga $\mathbf{Z}(\mathbf{t})$, em cada instante \mathbf{t} do período de previsão de duração \mathbf{T} , por uma série temporal da forma:

$$\{\mathbf{Z}(\mathbf{t}), \mathbf{t} = 1, 2, \dots, \mathbf{T}\} \quad (1)$$

Em sua forma mais simples, este modelo armazena os \mathbf{T} valores de carga baseados no comportamento prévio da carga observada. Algumas empresas utilizam o padrão da carga medida da semana anterior como um modelo para prever a carga da semana corrente.

Alternativamente, um conjunto de curvas é armazenado para semanas típicas do ano e para condições meteorológicas típicas, tais como, dias úmidos, secos e nublados, que são heurísticamente combinados com o padrão da carga semanal mais recente para realizar a previsão. Em tais casos, o julgamento do operador determina a previsão, sendo que fórmulas matemáticas explícitas não são utilizadas para descrever o modelo de carga.

Esta é uma área potencial de aplicação de sistemas especialistas que poderiam emular as regras fornecidas pelos operadores [5]. Uma das desvantagem desta abordagem é que a experiência é muitas vezes escassa e, conseqüentemente, a codificação do conhecimento pode ser limitada ou imprecisa. Logo, esta também é uma área em potencial para a aplicação de sistemas inteligentes híbridos [6] nos quais tanto o conhecimento do previsor quanto a utilização de técnicas avançadas de reconhecimento de padrões podem ser aplicados. Alguns trabalhos relativos à modelagem heurística foram realizados utilizando análise de grupos (cluster analysis) e técnicas clássicas de reconhecimento de padrões [7,8]. Um modelo hora-do-dia mais comum pode tomar a seguinte forma:

$$Z(t) = \sum_{i=1}^N \alpha_i f_i(t) + v(t), \quad t \in \tau \quad (2)$$

onde a carga no instante t , Z_t , é considerada como sendo uma soma de um número finito de funções periódicas $f_i(t)$, senos e cosenos de período 24 ou 168 hrs, dependendo do horizonte de previsão.

Os coeficientes α são considerados como parâmetros de dinâmica lenta e $v(t)$ representa o erro de modelagem, admitido ser um ruído branco. É admitido que o modelo seja válido para o intervalo de tempo τ que cobre os períodos de tempo do passado recente, do presente e do futuro, cobrindo o horizonte de interesse.

Os parâmetros α_i são estimados através da aplicação de regressão linear simples ou alisamento exponencial a um histórico de carga $\{Z_t, t \in \tau_{\text{past}}\}$ onde τ_{past} é um intervalo de tempo que cobre o passado recente da carga [9]. Exemplos de tais modelos podem ser encontrados nas referências [9,10].

A vantagem destes modelos é que eles são estruturalmente simples. Um lado negativo é que estes modelos não representam de forma precisa a correlação estocástica da dinâmica da carga. Além disso, uma vez que a adaptabilidade que estes modelos fornecem é baseada em parâmetros de ajuste arbitrário, que geralmente não se modificam após a calibração inicial, assim, as previsões obtidas são insatisfatórias para condições não usuais (por exemplo, entrada de frentes frias).

Existe uma segunda classe de modelos hora-do-dia, que são aqueles baseados em decomposição espectral que será tratado mais adiante. Os modelos dinâmicos reconhecem o fato de que a carga não é apenas uma função da variável tempo, mas, também função de entradas de informação meteorológica e de entradas aleatórias.

Os modelos dinâmicos são basicamente de dois tipos: modelos de Box & Jenkins e modelos de Espaço de Estado. Estes modelos serão abordados no capítulo 2.

1.7 Redes Neurais em previsão de carga

As redes neurais artificiais vem sendo aplicadas com sucesso em previsão de carga, tal interesse é motivo de pesquisa para muitos especialistas da área que procuram novos modelos ou arquiteturas que apresentem melhores resultados [11,12]. Muitos livros e artigos vêm sendo publicados, demonstrando o uso potencial desta nova técnica da Inteligência Artificial (IA). O capítulo 3 apresenta uma rápida revisão dos conceitos fundamentais das redes neurais.

1.8 Metodologia proposta

O objetivo deste trabalho é fazer um estudo comparativo dos modelos de previsão utilizando técnicas de Redes Neurais e métodos estatísticos clássicos como: AR, MA, ARIMA, etc. Também será objeto de estudo a influência do agrupamento, destendenciamento, sazonalização, tipos de treinamento, tamanho do conjunto de treinamento utilizados pela Rede Neural. Estes tópicos serão abordados no capítulo 4.

O capítulo 5 apresenta alguns resultados utilizando-se dados provenientes da Hydro-Quebec em sua barra Pouffe, na região da Grande Montreal.

CAPÍTULO 2

MÉTODOS DE PREVISÃO DA CARGA ELÉTRICA

2.1 Introdução

Diversas técnicas podem ser encontradas na literatura de previsão de carga a curto prazo, que usualmente refletem em previsões estatísticas. Vários algoritmos têm sido implementados, com diferentes graus de sucesso. Neste capítulo será abordado as diversas técnicas utilizadas em previsão de carga.

As técnicas de previsão de carga podem ser divididas em: quantitativas (ou estatísticas), aqui denominados de métodos convencionais; e qualitativas (ou baseadas em conhecimento), por exemplo: Sistemas Especialistas e Redes Neurais Artificiais.

2.2 Métodos quantitativos (estatísticos)

As técnicas quantitativas são baseadas em métodos matemáticos bem elaborados e comprovados [13]. Essas técnicas incluem: (1) métodos de regressão linear; (2) alisamento exponencial (exponential smoothing); (3) séries temporais; e (4) o método espaço-estado. Na revisão dessas técnicas é apresentada resumidamente nesta seção.

2.2.1 Método da Regressão Linear Múltipla (RLM)

No método da regressão linear múltipla, a carga é encontrada em termos de um conjunto de variáveis explicativas, tais como variáveis climáticas que influenciam a carga elétrica [14]. O modelo utilizado neste método é escrito, como:

$$y(t) = a_0 + a_1x_1(t) + \dots + a_nx_n(t) + a(t) \quad (1)$$

onde :

$y(t)$ = carga elétrica prevista

$x_1(t), \dots, x_n(t)$ = variáveis correspondentes a $y(t)$.

$a(t)$ = variável aleatória com média zero e variância constante

a_0, a_1, \dots, a_n = coeficientes da regressão

As variáveis explicativas do modelo são identificadas sobre as bases da análise de correlação de cada uma dessas variáveis (independente) com a variável (dependente) carga. Os coeficientes estimados da regressão são usualmente encontrados usando a técnica de estimação dos mínimos quadrados. Uma limitação para a utilização do método de regressão é que requer um número muito grande de dados para a estimação dos coeficientes do modelo.

2.2.2 Método de Alisamento Exponencial (MAE)

No item anterior abordou-se o método da regressão linear múltipla, que como foi visto relaciona a variável dependente $y(t)$ a um conjunto de variáveis independentes ou preditoras $x_n(t)$. Com esse modelo pode-se prever a variável dependente $y(t)$, dado que os valores de $x_n(t)$ são especificados. No método baseado no alisamento exponencial, a carga $y(t)$ no tempo t é modelada usando a seguinte equação [9]:

$$y(t) = \beta(t)^T f(t) + \varepsilon(t) \quad (2)$$

onde:

$f(t)$ = função do tempo

$\beta(t)$ = coeficientes desconhecidos

$\varepsilon(t)$ = erros não correlacionados

T = operador transposto

Os coeficientes estimados são encontrados diminuindo o erro médio quadrado para os N últimos intervalos da amostra.

2.2.3 Séries Temporais Estocásticas

Antes de abordar o método das séries temporais estocásticas é dada uma visão sucinta sobre séries temporais. Uma abordagem mais profunda do assunto pode ser encontrada em [15].

2.2.3.1 Conceitos Básicos das Séries Temporais

Uma série temporal é um conjunto de observações ordenadas no tempo [16]. O consumo mensal de energia em uma residência, o preço semanal de um produto, o valor de um índice de produção industrial são alguns exemplos.

Todos os métodos estatísticos de previsão de séries temporais baseiam-se na idéia de que as observações passadas da série contêm informações sobre o seu padrão de

comportamento no futuro. A essência desses métodos consiste em identificar o padrão da série, separando-o do ruído contido nas observações individuais, e utilizando-o para prever os valores futuros da série. Conceptualmente uma série temporal, é representada por um conjunto de observações $[X(t), t \in T]$; $t = 1, 2, \dots, T$ onde $X(t)$ é um conjunto de observações discretas, amostradas em tempos equidistantes e que apresentam uma "dependência serial" entre as mesmas, ou seja, são correlacionadas.

a. Objetivos

Existem dois objetivos principais na análise das séries temporais:

- 1) Entender o mecanismo do sistema gerador da série temporal e
- 2) Predizer o comportamento futuro do sistema.

Para isso, os problemas de interesse são:

- a) Descrever o comportamento da série. Neste caso, ferramentas úteis são: a construção de gráficos da série, construção de histogramas, obtenção de estatísticas descritivas simples, verificação de tendências, ciclos e variações sazonais
- b) Investigar o mecanismo gerador da série temporal
- c) Fazer previsões de valores futuros da série, a partir de valores passados
- d) Procurar periodicidades relevantes nos dados

b. Tipos de variação

A análise clássica de séries temporais trata, basicamente, da decomposição da série em três componentes: sazonal, cíclica, e tendência. O interesse por este enfoque vem sendo retomado recentemente [17]. Os diferentes tipos de variação que se procura analisar são:

Sazonal Muitas séries exibem uma variação periódica (diária, semanal, mensal, anual, etc). Este tipo de variação é fácil de entender, podendo ser medida e extraída dos dados, fornecendo uma série desazonalizada. Por exemplo: as cargas exibem uma periodicidade diária, refletindo níveis de atividades diferentes durante o dia, as medidas de temperatura exibem uma variação anual periódica, baixando durante o inverno e aumentando no verão.

Cíclica Além de efeitos sazonais, algumas séries exibem variações cíclicas. Por exemplo: modelos de curvas de cargas semanais refletem as mudanças das atividades industriais durante a semana. Nos fins de semana, as cargas são na média mais baixas que durante os dias de semana.

Tendência São os efeitos a longo prazo na média. O incremento do número de consumidores tende a um crescimento contínuo da carga, que somente poderia ser detectado durante longos períodos. A demanda por grupos de consumidores poderiam crescer a taxas diferentes durante o ano. Por exemplo, no verão a demanda da carga poderia crescer mais rapidamente devido ao incremento necessário para suprir equipamentos de refrigeração.

2.2.3.2 Método das séries temporais estocásticas (Box & Jenkins)

Este método parece ser o mais popular na previsão de carga. A teoria que sustenta estes modelos pode ser encontrada na referência [15]. Numa breve explicação, a série de carga $y(t)$ é modelada como sendo a saída de um filtro linear que possui como entrada uma série aleatória $a(t)$, usualmente chamada " ruído branco " como ilustra a figura 2.1. Esta entrada aleatória possui média zero e uma variância constante $\delta_a^2(t)$.

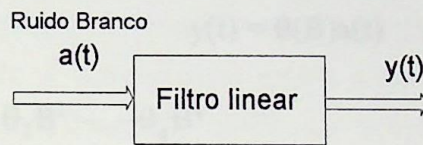


Fig. 2.1 - Modelagem da carga em séries temporais.

Dependendo da característica do filtro linear, diferentes modelos podem ser classificados como se segue nesta seção.

a. Modelo autoregressivo (AR)

No modelo autoregressivo, o valor da série temporal $y(t)$ é expresso linearmente em termos dos valores prévios ($y(t-1), y(t-2), \dots$) e um ruído aleatório $a(t)$. A ordem destes processos depende dos maiores valores prévios em que $y(t)$ é retroagido

Para um processo autoregressivo de ordem p , $AR(p)$, o modelo pode ser escrito como:

$$y(t) = \phi_1 y(t-1) + \phi_2 y(t-2) + \dots + \phi_p y(t-p) + a(t) \quad (3)$$

Introduzindo o operador \mathbf{B} que define $y(t-1) = \mathbf{B}y(t)$, e conseqüentemente $y(t-m) = \mathbf{B}^m y(t)$, a equação (3) pode ser escrita na forma :

$$\phi(\mathbf{B})y(t) = a(t) \quad (4)$$

onde: $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$

b. Modelo média - móvel (MA)

No modelo da média móvel, o valor da série temporal $y(t)$ é expressado linearmente em termos dos valores prévios e corrente de uma série de ruído branco $a(t), a(t-1), \dots$. Esta série de ruído é construído a partir dos erros previstos ou residuais quando observações nas cargas começam a estar disponíveis. A ordem do modelo depende do maior valor do ruído (erro) em que $a(t)$ é retroagido. Para uma média-móvel de ordem q , $MA(q)$, o referido modelo pode ser escrito como:

$$y(t) = a(t) - \theta_1 a(t-1) - \theta_2 a(t-2) - \dots - \theta_q a(t-q) \quad (5)$$

Uma aplicação similar do operador B sobre a série de ruído branco (erro), a equação (5) pode ser escrita como:

$$y(t) = \theta(B)a(t) \quad (6)$$

onde: $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$

c. Modelo da média - móvel autorregressiva (ARMA)

Este modelo é a combinação dos dois anteriores, ou seja, a combinação do modelo autorregressivo AR e do modelo da média móvel MA . No modelo média móvel autorregressivo, o valor da série temporal $y(t)$ é expresso linearmente em termos de seus valores prévios $y(t-1), y(t-2), \dots$ e em termos dos valores prévios e corrente de um ruído branco (erro) $(a(t), a(t-1), a(t-2), \dots)$. A ordem do modelo ARMA é selecionada pelo maior valor da série e pelo maior valor do ruído branco em que $a(t)$ é retroagido. Para um modelo média-móvel autorregressivo de ordem p e q , $ARMA(p,q)$, o modelo pode ser escrito como:

$$y(t) = \phi_1 y(t-1) + \dots + \phi_p y(t-p) + a(t) - \theta_1 a(t-1) - \dots - \theta_q a(t-q) \quad (7)$$

usando o operador B definido anteriormente, a equação (7) pode ser escrita da seguinte forma:

$$\phi(B)y(t) = \theta(B)a(t) \quad (8)$$

onde $\phi(B)$ e $\theta(B)$ foram definidos anteriormente.

d. Modelo média-móvel autoregressivo integrado (ARIMA)

O referido modelo foi introduzido por Box & Jenkins (1976). O modelo inclui parâmetros autorregressivos assim como parâmetros de média móvel, além de explicitamente incluir diferenciamento na formulação do modelo. Especificamente, os três tipos de parâmetros do modelo são: parâmetros autorregressivos (p), número de diferenciação (d), e parâmetros de média móvel (q). Na notação introduzida por Box & Jenkins, os modelos são representados como: ARIMA(p,d,q); por exemplo, um modelo representado como (0,1,2), possui zero (0) como parâmetro autorregressivo e dois (2) como parâmetro de média móvel. Estes parâmetros foram calculadas pela série e diferenciados uma vez (1).

Identificação

Como foi mencionado anteriormente, a entrada da série para o modelo ARIMA deve ser estacionária, isto é, deve possuir média e variância constantes através do tempo. Um simple teste de estacionaridade pode ser empregado através da série temporal $y(k)$, $k = 1, \dots, N$ utilizando as seguintes fórmulas:

$$(a) \text{ teste da média: } \bar{y}_i = \frac{1}{T} \sum_{k=i+1}^{i+T} y(k) \quad (9)$$

$$(b) \text{ teste da variância: } \sigma_i^2 = \frac{1}{T} \sum_{k=i+1}^{i+T} (y(k) - \bar{y}(k))^2 \quad (10)$$

Usualmente a série deve ser primeiramente diferenciada até esta ser estacionária. O número de vezes que a série precisa ser diferenciada para permitir estacionaridade é refletido no parâmetro d, exemplo: ARIMA(p,d,q). Para determinar o nível necessário de diferenciação, poderia ser examinada a curva de dados e o autocorrelograma.

Mudanças significativas no nível (mudanças para acima e para abaixo) usualmente requer diferenciação não sazonal de primeira ordem (atraso = 1); fortes mudanças de inclinação usualmente requer diferenciação não sazonal de segunda ordem. É importante destacar que algumas séries temporais poderiam ou não requerer ser diferenciada, e que séries sobre diferenciadas podem produzir coeficientes estimados menos estáveis.

Nesta fase que é usualmente chamado de "identificação". Nela necessita-se decidir quantos parâmetros autorregressivos (p) e média móvel (q) são necessários para produzir

um modelo efetivo parcimonioso do processo, ou seja, modelo que tem poucos parâmetros e maior número de liberdade entre os modelos.

Estimação e Previsão

O próximo passo é a estimação e a previsão. Os parâmetros são estimados usando procedimentos de minimização de funções. Esses parâmetros são utilizados para calcular novos valores da série, ou seja, a previsão (além dos dados de entradas incluídos) e intervalos de confiança para todos os valores previstos.

O processo de estimação é realizado com dados transformados (diferenciados). Uma vez que a previsão é gerada, a série precisa ser integrada, ou seja, fazer o inverso da diferenciação para que a previsão seja expresso em valores compatíveis com os dados de entrada.

e. Modelos Sazonais

Como resultado das periodicidades diárias, semanais, anuais e outras periodicidades, muitas séries temporais exibem comportamentos aleatórios em resposta a uma ou mais dessas periodicidades. Por conseguinte, diferentes classes de modelos que tem esta propriedade são designadas de Modelos Sazonais.

As séries temporais sazonais poderiam ser representadas como um modelo **AR**, **MA**, **ARMA**, ou como um modelo sazonal **ARIMA**. Têm sido demonstrado que o modelo geral multiplicativo $(p,d,q) \times (P,D,Q)_s$ para um modelo de série temporal pode ser escrito na forma [13]:

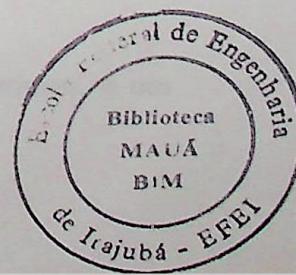
$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D y(t) = \theta(B)\Theta(B^s)a(t) \quad (11)$$

As expressões ∇_s^D , $\phi(B^s)$, e $\Theta(B^s)$ são obtidas como se segue:

$$\nabla_s^D = (y(t) - y(t-s))^D = (1 - B^s)^D y(t),$$

$$\phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps}, \text{ e}$$

$$\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_q B^{qs}$$



O modelo apresentado na equação (11) pode obviamente ser estendido para o caso onde duas sazonalidades são consideradas. A ordem do modelo é: $(p,d,q) \times (P,D,Q)_s \times (P',D',Q')_{s'}$ e é expresso como:

$$\phi(B)\Phi(B^s)\Phi'(B^{s'})\nabla^d\nabla_s^D\nabla_{s'}^{D'}y(t)=\theta(B)\Theta(B^s)\Theta'(B^{s'})a(t) \quad (12)$$

onde as expressões $\nabla^{D'}$, $\Phi'(B^{s'})$, e $\Theta'(B^{s'})$ são obtidas usando o segundo termo da série temporal sazonal como se segue:

$$\nabla_{s'}^{D'}=(y(t)-y(t-s'))^{D'}=(1-B^{s'})^{D'}y(t)$$

$$\Phi'(B^{s'})=1-\Phi'_1B^{s'}-\Phi'_2B^{2s'}-\dots-\Phi'_pB^{ps'} \quad , \quad e$$

$$\Theta'(B^{s'})=1-\Theta'_1B^{s'}-\Theta'_2B^{2s'}-\dots-\Theta'_qB^{qs'}$$

f. Método da função de transferência

Os modelos definidos anteriormente permitem que a função $y(t)$ ser expressa em termos dos seus valores históricos e um ruído branco. Se outras variáveis afetam o valor de $y(t)$, o efeito dessas variáveis pode ser considerada usando o modelo da função de transferência. Para o caso de uma variável independente $x(t)$, como a temperatura, o modelo por função de transferência ilustrado na figura 2.2 pode ser escrito na forma [13]:

$$y(t)=\frac{\omega(B)}{\sigma(B)}x(t-b)+n(t) \quad (13)$$

$$\omega(B)=\omega_0-\omega_1B-\omega_2B^2-\dots-\omega_rB^r$$

$$\sigma(B)=1-\sigma_1B-\sigma_2B^2-\dots-\sigma_sB^s$$

onde $\omega(B)$ e $\sigma(B)$ são polinômios em B , X_t é a variável de entrada da informação meteorológica, como por exemplo, a temperatura, e

b = atraso de tempo

$n(t)$ = ruído da série

As séries $n(t)$ podem ser modeladas em termos dos seus valores anteriores, e um ruído branco usando as técnicas previamente discutidas nas seções **a** a **e**.

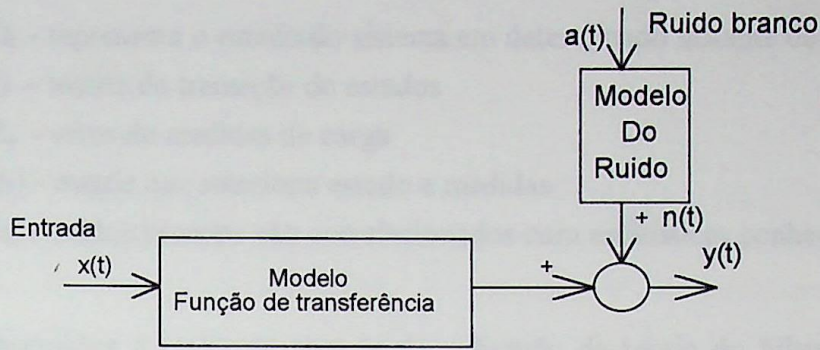


Fig. 2.2 - Modelo função de transferência.

A identificação do modelo da série é obtido pela análise dos dados de carga. Essa análise inclui, as curvas da função de autocorrelação (FAC) e a autocorrelação parcial (FACP). O uso dessas ferramentas leva a um cálculo inicial do grau de diferenciação para obter um modelo estacionário. As ordens dos polinômios do modelo autoregressivo (AR) e média móvel (MA) são inicialmente determinadas pelas funções da autocorrelação e da autocorrelação parcial.

Para o caso do modelo por função de transferência da série temporal, o gráfico da função da autocorrelação cruzada (FACC) entre a série temporal $y(t)$ (carga) e a série de entrada $x(t)$ (temperatura) é também necessário para a avaliação da resposta de atraso do tempo b e dos polinômios $\omega(B)$ e $\sigma(B)$ de ordem r e s . A estimação dos coeficientes do modelo de previsão de carga é feita usualmente através de um método de estimação eficiente.

2.2.4 Método de espaço de estado e filtro de Kalman (EEFK)

Este é um método geral de previsão que pode incluir os métodos previamente mencionados (por exemplo, modelos de Box & Jenkins) e mais, tais como modelos com coeficientes variando no tempo. Nesta metodologia, a carga é modelada como uma variável de estado utilizando formulação de espaço de estado que é definida por dois conjuntos de equações: um que descreve o comportamento dinâmico do estado do sistema (equações de estado) e outro correspondente a medições ou observações desse comportamento (equações de medição). Tais equações podem ser escritas para modelar o comportamento da carga como:

$$\text{Equações de Estado: } X(k+1) = \phi(k)X(k) + W(k) \quad (14)$$

$$\text{Equações de Medição: } Z(k) = H(k)X(k) + V(k) \quad (15)$$

onde: X_k - representa o estado do sistema em determinado instante de tempo t_k

$\phi(k)$ - matriz de transição de estados

Z_k - vetor de medidas de carga

$H(k)$ - matriz que relaciona estado e medidas

$W(k)$ e $V(k)$ - ruídos brancos não correlacionados com estatísticas conhecidas

A obtenção de previsões é realizada através da aplicação da teoria do Filtro de Kalman utilizando-se às equações (14) e (15) [13].

O método de Espaço de Estado é muito atraente para previsões on-line como resultado da propriedade recursiva do filtro de Kalman. A previsão ótima poderia ser baseado sobre o modelo assumido. Isto é, o modelo deve ser conhecido a priori para usar o filtro de Kalman. Em contrapartida, a identificação do modelo é muito laboriosa, e é a principal dificuldade do método. Isto se reflete especialmente na descrição estatística de $W(k)$ e $V(k)$, As matrizes de covariância dos ruídos são desconhecidas e não são facilmente estimadas.

2.3 Métodos qualitativos (baseados no conhecimento)

Devido ao crescimento dos sistemas de energia elétrica e, em consequência, do aumento da complexidade, uma série de fatores passaram a interferir no comportamento da carga. Com isto, o processo de previsão tem se tornado cada vez mais complexo e criado a necessidade de obter-se previsões com maior precisão. Entre esses fatores pode-se citar: intercâmbio de energia, gerenciamento da carga, tarifação horo-sazonal, etc. Surgiu então uma nova técnica que não esta baseada somente em análises matemáticas complexas, mas utiliza a experiência humana também. A esta técnica denomina-se Sistemas Especialistas.

2.3.1 Método baseado em Sistemas Especialistas

Os Sistemas Especialistas são novas técnicas aplicadas na previsão de carga, que surgiram como resultados dos avanços no campo da inteligência artificial nas últimas duas décadas. A razão principal para isto é que os sistemas especialistas são uma ferramenta poderosa para modelar eventos que são incertos e não exibem alguma regularidade [5,18].

Um sistema especialista é um programa computacional não algorítmico que pode conter um conjunto de regras que tem a habilidade de atuar como um especialista. Desta forma o programa pode "raciocinar" e resolver problemas [19]. O modelo de previsão de

carga é elaborado usando o conhecimento adquirido através da experiência de um especialista da área.

O conhecimento do especialista é armazenado, na forma de fatos e regras, em uma parte do sistema especialista chamado de base de conhecimento. Quando o sistema é posto a funcionar os dados externos provenientes de medidas e observações no sistema supervisionado são transformadas em fatos e escritos em uma área de trabalho (working memory). A manipulação das informações contidas na base de conhecimento e na área de trabalho é realizada pelo motor de inferência. É este motor que se responsabiliza por achar a resposta a um dado problema.

Embora o programa para sistemas especialistas possa ser escrito em praticamente qualquer linguagem (C, Pascal, Fortran), a grande maioria destes sistemas foi desenvolvido utilizando LISP (pelos americanos) e PROLOG (europeus e japoneses). A figura 2.3 ilustra a estrutura de um sistema especialista.

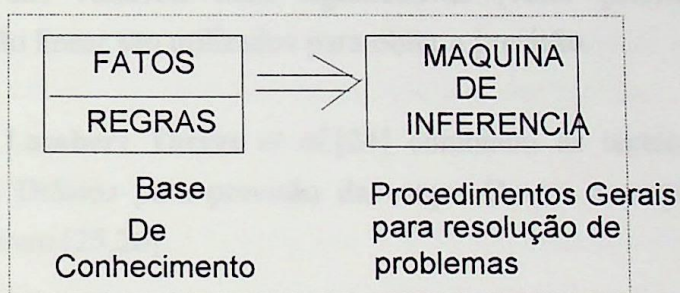


Fig. 2.3 - Estrutura de um sistema especialista.

2.3.2 Método baseado em Redes Neurais Artificiais (RNAs)

Pesquisas recentes na área da inteligência artificial têm encontrado uma alternativa para modelos de previsão que utilizam técnicas estatísticas clássicas: as redes neurais artificiais [20]. No método baseado em sistemas especialistas como foi abordado na seção anterior, quase sempre presume-se a existência de um especialista capaz de fazer a previsão com precisão. Já as Redes Neurais Artificiais não precisam da experiência ou da intervenção humana para tentar modelar a relação entre dados de entrada e saída observados [21,22].

A técnica de redes neurais não requer uma função que relacione dados passados de carga ou variáveis climáticas (temperatura, humidade e velocidade do vento) e cargas previstas. Isto é, possui a capacidade de se adaptar às mudanças na previsão através de

aprendizado próprio. Esta técnica utiliza casos de treinamento para prover ou mapear a relação entre as variáveis de entrada e saída, neste caso, a carga prevista.

2.4 Novas tendências

A nova tendência é fazer previsão usando técnicas híbridas, ou seja, uma combinação de várias técnicas de previsão como: redes neurais, sistemas especialistas e conjuntos difusos, por exemplo. No artigo de Rahman e Shresta [23] combinou-se com sucesso os métodos estatísticos e sistemas especialistas para a previsão da carga a curto prazo. A razão principal do método é reduzir a pesada carga computacional de técnicas estatísticas, assim como reduzir também grandes bancos de dados utilizados para testar modelos estatísticos. Em outra ordem, sistemas especialistas poderiam apresentar regras inconsistentes derivados de opiniões de especialistas.

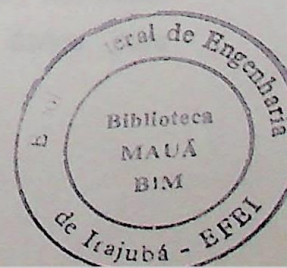
Para contornar este problema, os autores desenvolveram um método em que o sistema especialista contém regras de como extrair informação de um histórico de dados. Após a determinação das variáveis mais significativas (vetor prioritário), técnicas estatísticas como regressão linear são utilizados para obter a previsão.

O artigo de **G. Lambert Torres et al.**[24] combinou as técnicas de Sistemas Especialistas e Métodos Difusos para previsão da carga. Outros exemplos de sistemas híbridos podem ser vistos em [25,26].

2.5 Considerações práticas

Nesta seção são estabelecidas critérios a serem usados para o desenvolvimento de modelos de previsão de carga a curto-prazo. Também serão discutidos aspectos práticos na seleção e formulação de métodos de PCCP. Em geral, as seguintes etapas devem ser estabelecidas para a modelagem e previsão da carga.

- i) - Formulação ou seleção do modelo de previsão
- ii) - Identificação ou atualização dos parâmetros do modelo
- iii) - Avaliação do desempenho do modelo e atualização da previsão
- iv) - Retorno à etapa anterior, caso o desempenho seja satisfatório; ou caso contrário, retornar às etapas (i) ou (ii).



O desempenho do modelo deve ser continuamente monitorado. Após obter uma estrutura de modelo adequada, uma possível deterioração do desempenho do modelo deve ser corrigida inicialmente através do ajuste fino dos coeficientes do modelo (ii). Mudanças na estrutura do modelo, embora possíveis não são freqüentes, uma vez que um modelo adequado tenha sido escolhido. As previsões são atualizadas de acordo com o intervalo de discretização adotado, por exemplo, meia hora.

Os requisitos computacionais associados à função de previsão de carga são considerados modestos. Entretanto, para estudos que envolvam a previsão de carga por barra tais requisitos deverão crescer significativamente. Os operadores preferem recorrer a programas computacionais de previsão que sejam conversacionais e amigáveis, e que funcionem bem particularmente em períodos críticos, como por exemplo, na hora do pico de carga diária. A seguir, serão apresentadas as considerações práticas sobre as etapas usuais da metodologia geral de previsão citada anteriormente.

2.5.1 *Formulação e Seleção do Modelo*

A primeira consideração para a seleção de um modelo de previsão são os objetivos da previsão, como por exemplo: a natureza da grandeza a ser prevista, o horizonte de tempo desejado e o uso pretendido da previsão. Mais de um modelo pode ser necessário para prever o pico de carga diário do sistema, os valores da carga do sistema em períodos específicos do dia, os valores da carga horária (ou a cada 15 minutos) do sistema e/ou a energia semanal do sistema. Em certos casos, mais de um modelo pode ser usado para prever as mesmas grandezas através da combinação estatística das previsões [27].

Uma consideração importante na formulação e/ou seleção de modelos de previsão apropriados é a escolha de modelos parcimoniosos, ou seja, aqueles que tenham o menor número de variáveis explicativas (independentes) e de coeficientes de ajuste. Tais modelos são mais fáceis de se inicializar, atualizar e operar.

Por fim, os requisitos de dados necessários à modelagem também devem ser considerados. Tais requisitos estão intimamente ligados ao tipo de modelo formulado. Em geral, modelos que requerem um grande conjunto de dados para serem inicializados são os não-lineares, com muitos coeficientes e os que envolvem variáveis meteorológicas.

Em contraste, modelos cujos coeficientes apresentam uma relação linear, tais como os modelos hora-do-dia, usualmente requerem um número menor de dados para inicialização e atualização.

Deve-se lembrar também que existe um certo dilema envolvendo a massa de dados a se considerar. Por outro lado, é desejável que se procure estabelecer, tanto quanto possível, uma relação permanente entre a carga e suas variáveis explicativas, o que requer uma massa de dados que cubra um longo período de observação da carga.

Também, deseja-se que o modelo seja flexível o bastante para refletir mudanças de comportamento do processo em estudo, o que impõe a utilização de um conjunto de dados não muito extenso, cobrindo o período mais recente de carga. Em geral, modelos que utilizam uma menor massa de dados são preferíveis, tendo em vista sua facilidade de uso e atualização. Uma base de dados referente a períodos de 3 a 6 semanas formado a partir do dia presente é desejável, devendo tal base conter também informações sobre dias especiais que tem periodicidade anual. Para efeitos de inicialização dos modelos convencionais, um conjunto de dados de carga e fatores meteorológicos de 2 a 3 anos, observados em base horária, é suficiente.

Considerando ainda a fase de inicialização, o modelo de previsão deve ser capaz de identificar e tratar dias especiais, tais como, feriados, início e término de período escolar, etc. Além disso, antes de prosseguir com a fase de identificação, a carga deve ser examinada para detectar qualquer comportamento anormal o que pode ser causado por eventos tais como greves, blackouts, dias de eleição, ou programas especiais de T.V.

Considerando o estado da arte atual da modelagem e previsão de carga é razoável afirmar que deve-se procurar desenvolver modelos que sejam capazes de considerar a forma da curva de carga, bem como, a hora do dia, aspectos climáticos, efeitos aleatórios, e acompanhar a dinâmica da carga. Modelos que descrevem apenas o pico de carga, não envolvendo efeitos meteorológicos, embora de mais simples desenvolvimento, não oferecem a precisão e a flexibilidade dos modelos mais formais.

2.5.2 *Filtragem dos dados*

Antes da aplicação das técnicas de identificação, deve-se levar em conta tanto as variações sazonais da carga, quanto o possível crescimento ou o declínio das tendências de um ano para outro. Uma forma de manipular esta variação é pré-filtrar os dados com um período de um ano, com isso eliminando as variações sazonais do processo de carga (pré-filtrado) que é então suposto estacionário.

2.5.3 Avaliação de desempenho

O desempenho de um método de previsão de carga pode ser avaliado em termos de precisão do modelo adotado, da facilidade do uso do programa computacional desenvolvido e pela capacidade de detecção e correção de condições anormais nos dados utilizados.

A avaliação da precisão de um modelo requer que o erro de previsão, isto é, a diferença entre os valores previsto de carga e o “medido”, seja obtido para cada intervalo de tempo do período de previsão. Na prática, para horizonte de previsão de 24 horas, um modelo é considerado preciso se apresentar erros de previsão menores que 3%.

Dois principais fatores atuam como limitadores da precisão dos modelos de previsão: o horizonte da previsão e as incertezas das variáveis explicativas do comportamento da carga. A medida que o horizonte de tempo aumenta, a precisão da previsão se deteriora. Também, quanto maior o número de variáveis explicativas no modelo, mais incerteza é introduzida na previsão. Isto é particularmente verdade quando as variáveis explicativas baseadas em previsão possuem uma grande incerteza associada.

Além disso, deve-se estar consciente que as diferentes variáveis meteorológicas previstas possuem diferentes precisões. Por exemplo, é consideravelmente mais fácil prever a temperatura do que prever o nível de precipitação.

Conseqüentemente, a utilização de variáveis independentes que são difíceis de prever devem ser evitadas de modo a prevenir a possibilidade de gerar previsões de carga com grandes erros envolvidos.

A comparação da precisão de dois ou mais modelos diferentes de PCCP deve ser avaliada sob condições que se aproximem o mais próximo possível das condições de operação verdadeiras. Para que a comparação tenha sentido, a avaliação deve ser realizada em um grande número de sub-períodos de um período suficientemente longo. Esta abordagem permite que o desempenho de diferentes modelos possa ser comparado em uma base uniforme sobre um amplo espectro de conjunto de dados.

Testes para a avaliação de desempenho devem ser realizados sistematicamente pela verificação dos erros de previsão um passo a frente, que devem constituir um processo do tipo ruído branco. Uma deterioração do modelo devido às variações dos parâmetros (coeficientes do modelo) ou por comportamento anormal da carga pode ser sistematicamente detectados pelo teste de ruído branco. Dependendo do tipo de deterioração, os coeficientes dos modelos devem ser atualizados, ou no caso de dados com

erros grosseiros ou anormais, tais dados podem ser descartados para obter-se um conjunto de dados “limpo”.

Outro ponto que merece destaque na avaliação de métodos de previsão diz respeito à facilidade de uso do programa computacional implementado. Este deve ser projetado de modo a incluir: acesso direto às variáveis meteorológicas, interface homem-máquina amigável e capacidade de rejeição de dados espúrios. Devem ser evitados modelos que necessitem, de excessiva quantidade de dados de entrada, atualização periódica de muitos coeficientes e ajuste freqüente de parâmetros.

Uma característica extremamente importante de um programa computacional de previsão de carga é a habilidade de detectar e excluir dados com erros grosseiros ou referentes a comportamento anormal da carga e prover substituição com valores corretos. Por exemplo, todo programa computacional de previsão deverá ter a flexibilidade de entrada manual de dados. Os despachantes, no curso de suas atividades, irão de tempo em tempo cometer erros na entrada de dados. O programa computacional de previsão deve ser “inteligente” o bastante para detectar e excluir tais erros e requisitar dos despachantes valores corretos.

Um assunto mais complexo é aquele de dados anormais. Uma hipótese implícita de todos os modelos de PCCP é que a carga está essencialmente em um modo de comportamento de estado estacionário. A ocorrência de feriados e dias especiais, entretanto, viola esta hipótese. Tais dados devem ser cuidadosamente identificados e tratados de forma a não interferirem nas previsões realizadas em dias subseqüentes considerados como normais.

CAPÍTULO 3

REDES NEURAIIS ARTIFICIAIS

3.1 Introdução

Desde o advento do primeiro computador digital em 1946 até o final da década de 1980, praticamente todas as aplicações para o processamento de dados e informações adotaram uma única abordagem básica: a "computação programada". Ela consiste no desenvolvimento prévio de um algoritmo, ou seja, um conjunto detalhado e rígido de instruções previamente estabelecidas para a resolução do problema em questão, o qual era codificado em alguma linguagem de computação.

Isto implicou no fato de que tal abordagem somente poderia ser utilizada em casos onde o processamento a ser efetuado pudesse ser descrito em termos de um conjunto de regras conhecido. Contudo, nem sempre isso acontece; muitas vezes a dedução desse conjunto de regras pode ser difícil. Além disso uma vez que os computadores atuais trabalham de forma totalmente lógica, o programa final tem de estar praticamente perfeito para poder funcionar.

Logo, o desenvolvimento de programas para computador é, na verdade, uma sucessão de ciclos: projeto, teste, e melhoria, que pode vir a demandar muito tempo, esforço e dinheiro. No final da década de 1980 surgiu uma abordagem revolucionária para o processamento de dados e informações: as Redes Neurais Artificiais, também conhecida como "perceptrons". Ela não requer o desenvolvimento de algoritmos ou conjunto de regras para analisar os dados, o que freqüentemente reduz de forma significativa o trabalho de desenvolvimento de programas que uma dada aplicação venha requerer. Na maior parte dos casos, a rede neural passa por um processo de treinamento a partir de casos reais conhecidos, adquirindo a partir daí a sistemática necessária para executar adequadamente o processamento desejado dos dados fornecidos. Ou seja: ela tem a capacidade de extrair conhecimento a partir de dados reais, dispensando qualquer modelo prévio já conhecido. Esta é a abordagem utilizada nos sistemas nervosos biológicos, particularmente em seres humanos [28].

Com efeito, pode-se perceber melhor a diferença entre a computação programada e as Redes Neurais Artificiais comparando-se computadores e seres humanos. Por exemplo, um computador efetua operações matemáticas com rapidez e precisão muito superiores aos seres humanos. Em contrapartida, estes conseguem reconhecer faces e imagens complexas de maneira muito mais precisa, eficiente e rápida que o melhor computador disponível atualmente.

Uma das razões dessa diferença de desempenho em tarefas diversas pode estar na forma como se organizam computadores e sistemas nervosos. Geralmente um computador consiste de um processador trabalhando sozinho, executando instruções fornecidas por um programador, uma a uma. Já sistemas nervosos consistem de bilhões de células nervosas, ou seja, neurônios com alto grau de inter conexão entre elas, que efetuam cálculos simples sem que haja a necessidade de que sejam programadas [28].

Logo, as redes neurais artificiais são um modo de se simular e tentar entender o que se passa nos sistemas nervosos biológicos, na esperança de se conseguir tomar proveito dos poderosos recursos desses sistemas orgânicos. Esta técnica revolucionária de inteligência artificial, apesar de ainda se encontrar em pleno desenvolvimento, já se encontra num estágio suficientemente adiantado para ser bastante útil em muitas aplicações. O desenvolvimento de micro computadores cada vez mais poderosos, estações de trabalhos, e processamentos paralelos tornou essa técnica plenamente acessível.

3.2 Fundamento das redes neurais artificiais

A técnica das redes neurais artificiais consistem na emulação de sistemas nervosos biológicos em programas ou circuitos digitais. Tais redes são capazes de estabelecer relações complexas entre dados de uma forma relativamente precisa, sem que seja necessário informar a elas qual a função que os relaciona. A Fig. 3.1 ilustra o neurônio biológico.

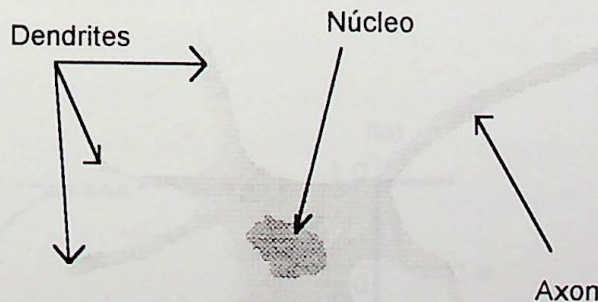


Fig. 3.1 - Neurônio Biológico

O elemento básico de uma rede neural artificial chama-se, **neurônio**, conhecido também por **nó** ou **elemento processador**, como pode ser visto na Figura 3.2.

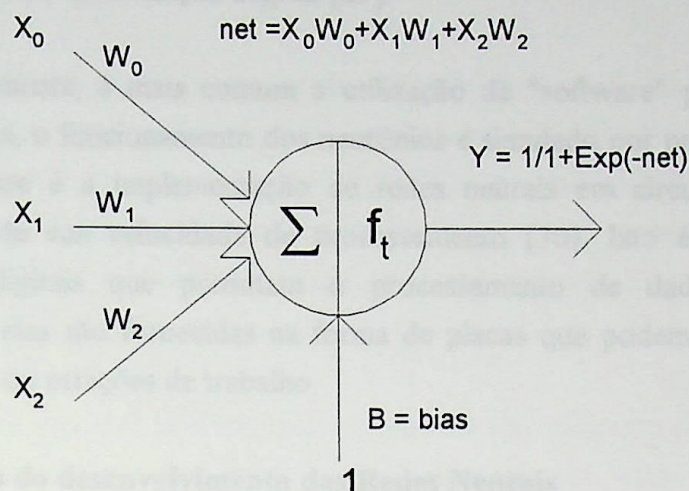


Figura 3.2 - Elemento básico de uma rede neural

Os neurônios da rede estão dispostos em camadas. Geralmente estão ligados a outros neurônios, em camadas anteriores e posteriores. Essas ligações são chamadas de sinapses (pesos). Pode haver ligações entre neurônios de uma mesma camada ou mesmo entre eles próprios.

Os sinais que são fornecidos a um dado neurônio são, na verdade o estado ou o valor de ativação dos neurônios precedentes, os quais são multiplicados por um peso correspondente. O estado, ou valor de ativação, do neurônio em questão, é calculado a partir da aplicação de uma função de limiar ao valor de entrada fornecido ao neurônio, ou seja, a somatória dos valores de ativação dos neurônios precedentes, multiplicados pelos respectivos pesos. A função de limiar (ou de ativação) é um tipo de função não-linear, no qual garante a plena funcionalidade das redes neurais com múltiplas camadas. Na figura 3.3 aparecem algumas das funções de ativação mais utilizadas, a escolha depende da aplicação a ser desenvolvida.

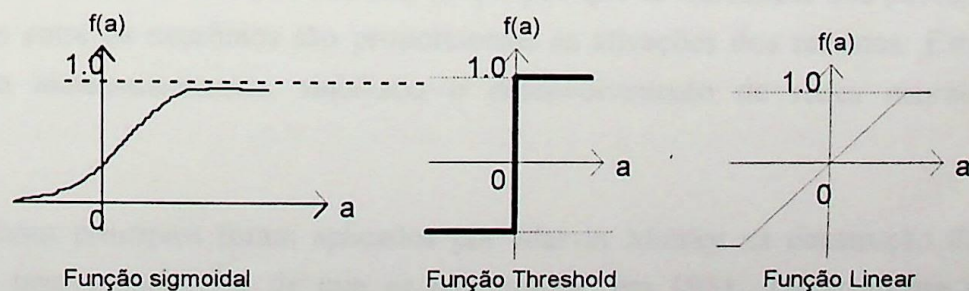


Fig. 3.3 - Funções de ativação mais utilizadas

No caso de redes neurais que processam dados analógicos utilizam uma função sigmoideal. Já para redes neurais cujos neurônios assumem apenas dois valores: zero (0) ou um (1) ela deve ser uma função degrau [29].

Atualmente, é mais comum a utilização de "software" para a geração de redes neurais, ou seja, o funcionamento dos neurônios é simulado por programas. Uma tendência bastante recente é a implementação de redes neurais em circuitos digitais, visando à maximização de sua velocidade de processamento [30]. Isto é conseguido através de arquiteturas digitais que permitam o processamento de dados de forma paralela. Normalmente elas são fornecidas na forma de placas que podem ser acopladas a micro-computadores ou estações de trabalho.

3.3 Histórico do desenvolvimento das Redes Neurais

Uma vez que o conceito e as características atuais das redes neurais definidos anteriormente, é interessante verificar como ocorreu a concepção e aperfeiçoamento desta técnica. Entre o final do século XIX e o início do século XX, pesquisadores como Sigmund Freud e William James estabeleceram as bases do funcionamento da neurologia dos seres vivos e, em particular, do ser humano.

Em 1943, Warren Mc Cullough e Walter Pitts estabeleceram as bases das Redes Neurais, concebendo procedimentos matemáticos análogos ao funcionamento dos neurônios biológicos. Este desenvolvimento foi puramente conceptual, uma vez que esses autores não sugeriram aplicações práticas a partir de seu trabalho, mesmo porque os sistemas propostos por eles não tinham capacidade de aprendizado. Note-se que o desenvolvimento de computadores naquela época era extremamente rudimentar.

Entretanto, através de cálculos manuais, foram propostos modelos muito simplificados de redes neurais artificiais, os quais podiam efetuar operações aritméticas ou lógicas. Em 1949, Donald Hebb sugeriu um modo de se proporcionar capacidade de aprendizado às redes neurais artificiais, ele propôs que as mudanças dos pesos, ou seja, as conexões entre os neurônios são proporcionais às ativações dos mesmos. Este princípio, traduzido matematicamente, viabilizou o desenvolvimento de redes neurais artificiais eficazes.

Esses princípios foram aplicados por Marvin Minsky na construção do Snark, o primeiro neurocomputador de que se tem notícia, em 1951. Tecnicamente ele foi um sucesso, uma vez que ele ajustava automaticamente os pesos entre as diversas sinapses, ou

seja, demonstrava, ao menos teoricamente, que tinha capacidade de aprendizado. Contudo, este dispositivo nunca executou qualquer função útil no campo do processamento de informações, constituindo-se portanto numa curiosidade acadêmica.

Em 1957, Frank Roseblatt concebeu um dispositivo denominado de *perceptron*, ou seja, uma rede neural automatizada, com duas camadas de neurônios, capaz de aprender de acordo com as regras propostas por Hebb. Tal aparelho, o Mark I, tinha capacidade de treinamento supervisionado e foi utilizado com sucesso no reconhecimento de caracteres. Pela primeira vez, as redes neurais artificiais foram utilizadas com sucesso numa aplicação prática.

Em 1962, Bernard Widrow desenvolveu um tipo diferente de processador para redes neurais, denominado Adaline [31], o qual dispunha de uma poderosa estratégia de aprendizado. Contudo, após esses desenvolvimentos, as redes neurais entrou numa grande crise. O perceptron de Roseblatt, que dispunha de apenas duas camadas de neurônios, era incapaz de realizar funções mais complexas que o reconhecimento de caracteres. Ele propôs como solução aumentar o número de camadas, mas apesar de toda sua visão e perspicácia neste campo, não logrou desenvolver um método de aprendizado eficaz para essas redes neurais mais avançadas. A verdade é que o desenvolvimento das redes neurais artificiais até aquele momento tinha ocorrido de forma predominantemente empírica. Havia ainda uma enorme carência de embasamento teórico e matemático formal para que sua evolução pudesse continuar de forma segura.

Este impasse incentivou o início de uma campanha orquestrada por grupos de pesquisa rivais para se suprimir as verbas concedidas ao desenvolvimento das redes neurais. Em 1967 ela logrou sucesso: as linhas de financiamento originalmente atribuídas ao estudo das redes neurais foram realocadas para outras áreas de pesquisa dentro do campo da Inteligência Artificial.

Apesar do descrédito gerado sobre a área das redes neurais, entre 1967 e 1982 os estudos neste campo continuaram, ainda que englobadas em outras linhas de pesquisa, como processamento adaptativo de sinais, reconhecimento de padrões e modelamento biológico. Este trabalho, ainda que silencioso, construiu as bases necessárias para que o desenvolvimento das redes neurais pudesse continuar de forma consistente.

Em 1974, Paul Werbos conseguiu o maior progresso em termos de redes neurais desde o perceptron de Roseblatt, ele lançou as bases do algoritmo de retropropagação ("back-propagation"), que permitiu que redes neurais com múltiplas camadas apresentassem

capacidade de aprendizado. Em 1982, David Parker desenvolveu um método similar, de forma aparentemente independente. Contudo, a potencialidade desse método tardou a ser reconhecida.

O desenvolvimento de uma base teórica sólida sobre as redes neurais artificiais e o vertiginoso desenvolvimento dos micro-computadores digitais fizeram retomar os estudos neste campo. Os primeiros resultados da retomada do desenvolvimento sobre redes neurais foram publicadas em 1986 e 1987, onde ficou consagrada a técnica de treinamento por retro propagação [32]. Nesta fase destacaram-se o surgimento do treinamento não-supervisionado, proposto por Teuvo Kohonen em 1984 [33] e as novas topologias de redes neurais, como a proposta por Bernard Kosko em 1987 (B.A.M.: "Bi-Directional Associative Memory").

Todo esse desenvolvimento resultou no advento de programas comerciais para micro computadores e estações de trabalho para o desenvolvimento e implementação de redes neurais com diversas topologias. Isso levou à popularização desta técnica e, de fato, ela está sendo crescentemente aplicada nos mais diversos campos do conhecimento humano. Contudo, a pesquisa sobre redes neurais artificiais continua em ritmo acelerado, uma vez que ainda há uma grande carência de formalismo matemático para que se possa explicar consistentemente seu modo de operação. Além disso, o nível de complexidade das novas topologias vem aumentando dia a dia, o que promete seu uso em aplicações cada vez mais intrincadas e desempenho cada vez melhor.

De toda esta evolução há alguns pontos que devem ser destacados. É interessante notar que tópicos aparentemente moderníssimos, como redes neurais e reconhecimento de caracteres, na verdade vem sendo pesquisados já há várias décadas. Pode-se também perguntar por que a computação programada se desenvolveu mais rapidamente que a neurocomputação, uma vez que ambas foram concebidas na mesma época. Isto pode ser explicado pelo fato de que as redes neurais, com sua multiplicidade de sinapses e complicados algoritmos de aprendizado, requerem computadores de alto desempenho para que possam ser desenvolvidas, equipamentos esses que só se tornaram facilmente disponíveis a partir do final da década de 1980, com o advento da tecnologia VLSI para a fabricação de circuitos para computador. Já a computação programada apresentou desempenho muito bom desde o surgimento de máquinas rudimentares como o ENIAC, em 1946.

3.4 Características do desempenho das Redes Neurais

Todo o "conhecimento" de uma rede neural está armazenado em suas sinapses, ou seja, nos pesos relativos às conexões entre os neurônios. Ele é adquirido por um processo de treinamento, que consiste na apresentação de seqüência de dados conhecidos à rede neural, gerando um processo de ajuste dos pesos das sinapses de forma a capturar o "conhecimento". O treinamento pode ser supervisionado ou não. No primeiro caso, é concluída no processo de treinamento uma etapa de verificação dos resultados calculados pela rede neural em treinamento. O erro observado a partir da comparação de resultados reais com os calculados pode ser utilizados para ajustar os pesos das sinapses de forma a aumentar a precisão da resposta da rede. Nas redes neurais onde o treinamento não é supervisionado ocorre um processo interno de organização dos dados que leva a um grau ótimo de ajuste de forma totalmente automática.

De fato, as redes neurais não prescindem totalmente da computação programada. Afinal, em qualquer aplicação, a aquisição dos dados, sua formatação e a saída dos resultados ainda utiliza esse recurso. Redes neurais são geralmente tratadas como subrotinas que podem ser amalgamadas com programas onde sua capacidade se faz necessária [29]. De modo geral, o uso de redes neurais apresenta as seguintes vantagens [34]:

- Há menor necessidade de se determinar a priori quais são os fatores determinantes sobre o modelo que está sendo desenvolvido
- É permitida a interferência de múltiplos fatores de entrada (ou seja, múltiplas variáveis), permitindo um inter-relacionamento muito mais complexo entre elas
- Alta tolerância a falhas, uma vez que é permitida a entrada de grande número de parâmetros. Tolerância a falhas refere-se ao fato que em muitas redes neurais, se alguns neurônios são destruídos, desabilitados, ou suas conexões são ligeiramente alterados, o comportamento da rede é ligeiramente degradado, ou seja o sistema não para abruptamente
- Modelamento direto do problema, sem a necessidade de se seguir um modelo preestabelecido, como no caso da regressão estatística
- Paralelismo inerente: cada sinapse na rede neural pode ser seu próprio processador

No presente trabalho serão considerados em detalhe apenas as redes neurais do tipo Rummelhart, ou seja, unidimensionais, com fluxo unidirecional de dados e treinamento supervisionado. No momento esta é a rede neural mais utilizada nos diversos campos da tecnologia, inclusive em previsão de carga e controle de processos.

De fato, certas características das redes neurais, como tolerância a falhas, robustez e capacidade de implementar uma classe particular de transformações, são garantidas por teoremas matemáticos. Ou seja, eles asseguram que as redes neurais podem ser empregadas de forma útil e confiáveis.

Por outro lado, esses teoremas nada afirmam sobre como (em termos conceitualmente mais altos) a rede neural aprende o "conhecimento". Suspeita-se que a descoberta desse mecanismo requeira uma verdadeira revolução intelectual na área do processamento de informações. Essa falta de embasamento teórico das redes neurais ainda é uma séria desvantagem desta técnica, uma vez que gera alguma desconfiança por parte dos especialistas quanto a sua confiabilidade. De fato, como já foi visto, no passado tal fato levou a uma virtual paralisação no desenvolvimento desta área devido a problemas matemáticos aparentemente insolúvel. Note-se, contudo, que atualmente a utilização de redes neurais está consolidada em diversas áreas, a pesquisa básica nesse campo é febril e os avanços teóricos são animadoras [32].

Uma outra grande desvantagem das redes neurais é o tempo requerido em sua fase de aprendizado, particularmente nas do tipo Rummelhart, que utilizam o método da *retropropagação*. Quanto mais sutis as relações entre as variáveis, e maior a precisão requerida nos resultados, maior será o tempo de treinamento. Em alguns casos críticos, poderão ser necessários dias de treinamento, mesmo utilizando-se micro computadores tão avançados. Mas mesmo este problema vem sendo resolvido, e em duas frentes:

- Através de algoritmo de treinamento mais rápidos e
- Através da utilização de arquitetura digital paralela nos circuitos digitais.

3.5 Modelos de Redes Neurais

Os modelos de redes neurais são especificados pelas topologias da rede, pela características dos nodos, em função do fluxo de dados (propagação para frente "feed-forward networks" ou nos dois sentidos "feedback network"), em função da ausência ou não de supervisão durante o treinamento da rede neural, e pelo treinamento ou regras de aprendizado [29].

Essas regras especificam os pesos iniciais fixados e como esses pesos poderiam ser adaptados durante o treinamento para melhorar a performance da rede. Arquiteturas e regras de treinamentos são tópicos de muitas pesquisas na atualidade. A continuação será citado algumas redes mais conhecidas.

3.5.1 Modelo Multi Camadas (feed-forward Network)

O modelo da rede multi camadas (unidirecional), também conhecido como Rumelhart [29] pode ser visto na figura 3.4. Este tipo de rede consiste de uma camada de entrada, uma ou mais camadas escondidas, e uma camada de saída. Cada camada está formada por um conjunto de neurônios e cada neurônio em uma camada está ligado aos neurônios na camada adjacente com diferentes pesos (w). Os sinais fluem da camada de entrada, passam pelas camadas escondidas, e chegam nas camadas de saídas. Com a exceção das camadas de entrada, cada neurônio recebe sinal dos neurônios das camadas prévias com respectivos pesos. O neurônio então produz um sinal de saída pelo passo dos sinais somados através de uma função sigmoideal.

Este tipo de rede é muito utilizado devido a sua simplicidade e estabilidade, sendo aplicadas para classificação, análise e interpolação de dados, previsão de carga etc. O comportamento deste tipo de rede é estático; ela se comporta de modo a tomar sua saída um reflexo da respectiva entrada. Ela deve ser treinada de modo a produzir os resultados desejados a partir da apresentação de dados reais.

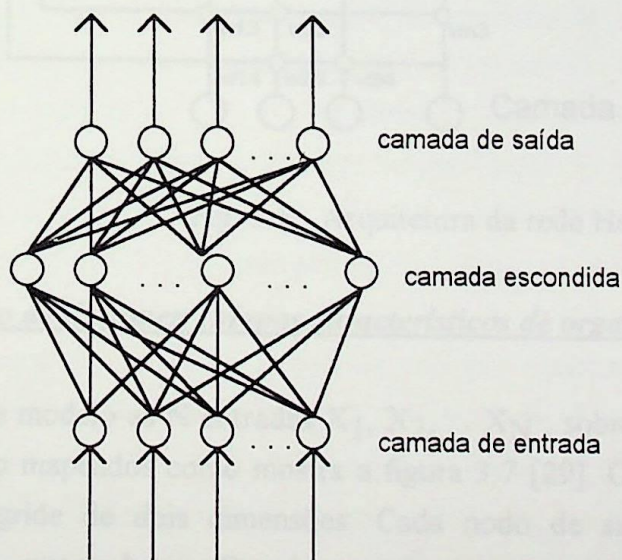
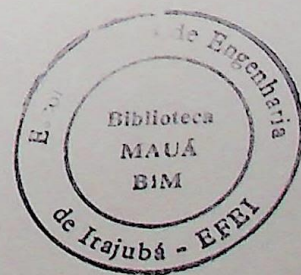


Fig. 3.4 - Rede multicamadas



3.5.2 Modelo Hopfield

A rede clássica tipo hopfield [35] utiliza três camadas: Uma camada de entrada, uma camada "Hopfield", e uma camada de saída. Cada camada possui os mesmos números de neurônios. Nesta rede há total integração entre os dados de entrada e os resultados obtidos, pois todos os neurônios são ligados entre si.

Este modelo se caracteriza por apresentar fluxo de dados multidirecional. Seu comportamento é dinâmico, mais complexo que o das redes de Rummelhart. Note-se que neste caso não há camadas discretas de neurônios. Na figura 3.5 pode ser visto a rede tipo Hopfield.

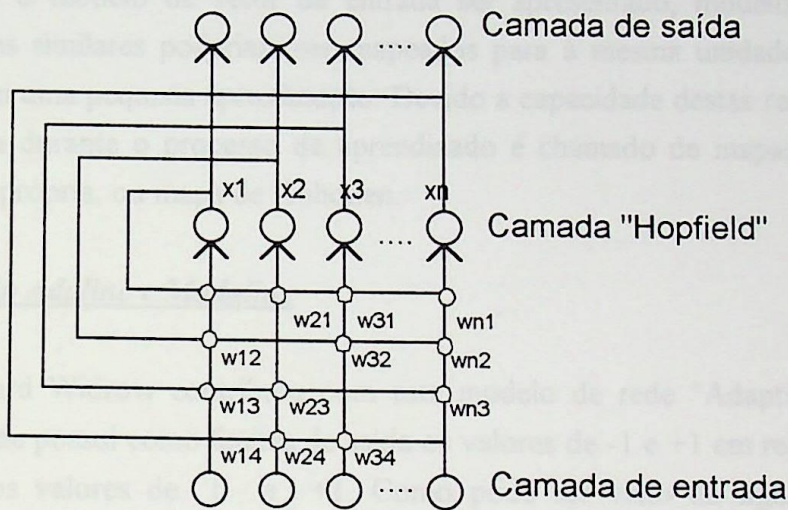


Fig. 3.5 - Arquitetura da rede Hopfield

3.5.3 Modelo de Kohonen (Mapas característicos de organização própria)

Neste modelo as N entradas X_1, X_2, \dots, X_N , sobre os M nodos de saídas Y_1, Y_2, \dots, Y_M são mapeados como mostra a figura 3.7 [29]. Os nodos de saída são adaptados sobre uma gride de dois dimensões. Cada nodo de saída é descrito por um par de coordenadas retangulares. O número de nodos de saída da gride (M) pode ser arbitrariamente escolhido. Como mostra a figura, cada unidade de entrada i é conectada às unidades j através das conexões dos pesos W_{ij} . Quando um modelo de entrada é apresentado para a rede neural sem especificar as saídas desejadas, a rede calcula os valores de ativação para cada nodo de saída baseado na presente conexão dos pesos.

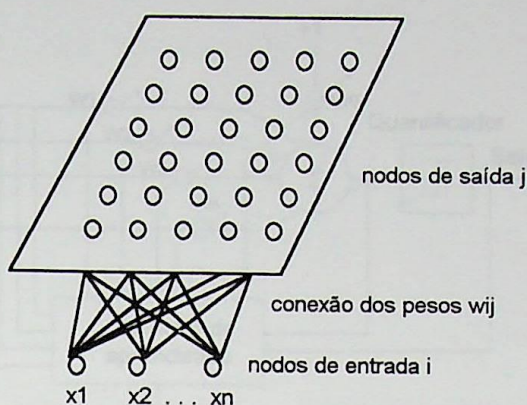


Fig. 3.6 - Estrutura do mapa característico de organização própria

Após o modelo de vetor de entrada ser apresentado, modelos de entrada com características similares poderiam ser mapeados para a mesma unidade de saída ou para unidades com uma pequena aproximação. Devido a capacidade destas redes de se organizar propriamente durante o processo de aprendizado é chamado de mapas característicos de organização própria, ou mapa de Kohonen.

3.5.4 Modelo Adaline e Madaline

Bernard Widrow contribuiu com este modelo de rede "Adaptive Linear Neuron" [35]. Esta rede possui como limites de saída os valores de -1 e +1 em relação ao perceptron que utiliza os valores de 0 e +1. Como pode ser visto na figura 3.7, o elemento processador básico do Adaline consiste de uma série de ajustes de pesos. O peso W_0 ou "bias" é ligado a uma entrada fixa de valor +1. As entradas são de bi-estados, ou seja podem ter valores entre +1 e -1. Os pesos podem ter valores positivos e negativos.

A soma de todos os pesos na entrada, incluindo o bias são aplicados a um quantizador que converte as saídas para os valores de -1 e +1. Através de um algoritmo de aprendizado os pesos da rede são ajustados. Na figura 3.8 é ilustrado a arquitetura do modelo Madaline. Este modelo não é nada mais que um conjunto de adalines em paralelo com uma saída que toma um voto majoritário sob as entradas. Este tipo de rede poderia ser aplicado em várias áreas como: previsão do tempo, e reconhecimento da voz.

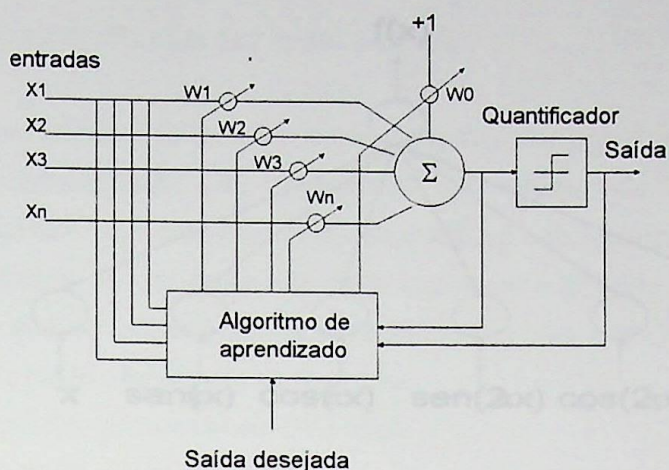


Fig. 3.7 - Elemento processador do Adaline

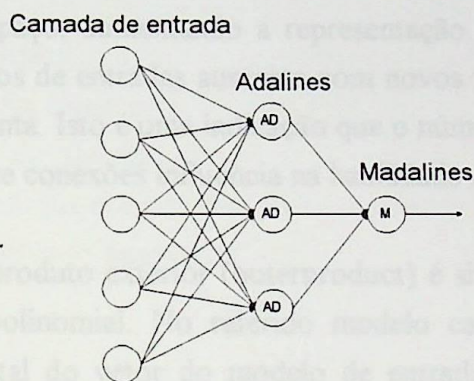


Fig. 3.8 - Arquitetura da rede Madaline

3.5.5 Modelo Link-Net

O modelo da expansão funcional "Link-Net" [34] é uma rede feed-forward que usa algoritmo standard back-propagation para ajustar os pesos. O referido modelo é diferente do modelo standard back-propagation, devido à introdução de nodos adicionais na camada de entrada a fim de melhorar a taxa de aprendizado. Dois modelos são considerados, o modelo da expansão funcional, e o modelo produto, respectivamente.

No modelo da expansão funcional, o Link funcional atua em cada nodo separadamente. Por exemplo, em adição para X_i , os nodos de entrada adicionais são forçados para ser $\sin\pi x$, $\cos\pi x$, $\sin 2\pi x$, $\cos 2\pi x$, e assim por diante, como pode ser visto na fig.3.9.

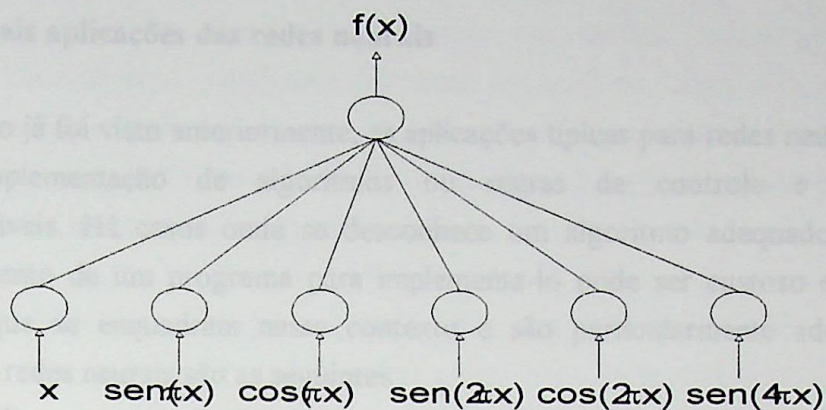


Fig. 3.9 - Modelo da expansão funcional

As funções seleccionadas forçam a ser um passo de funções ortogonais tais como funções seno e coseno de $n\pi X_i$. O efeito deste modelo é mapear os vetores de entrada num grande modelo espaço, aumentando a representação. Pode ser também notado que como o número de nodos de entradas aumenta com novos termos funcionais, o número de conexões também aumenta. Isto é uma indicação que o número de exemplos de treinamento em relação ao número de conexões influencia na habilidade de uma rede para generalizar.

No modelo do produto exterior (outerproduct) é simplesmente um caso especial do modelo da expansão polinomial. No referido modelo cada componente do modelo de entrada multiplica o total do vetor do modelo de entrada. O Link funcional neste caso genera um vetor de entrada para cada componente individual. Na figura 3.10 pode ser visto o modelo do produto exterior.

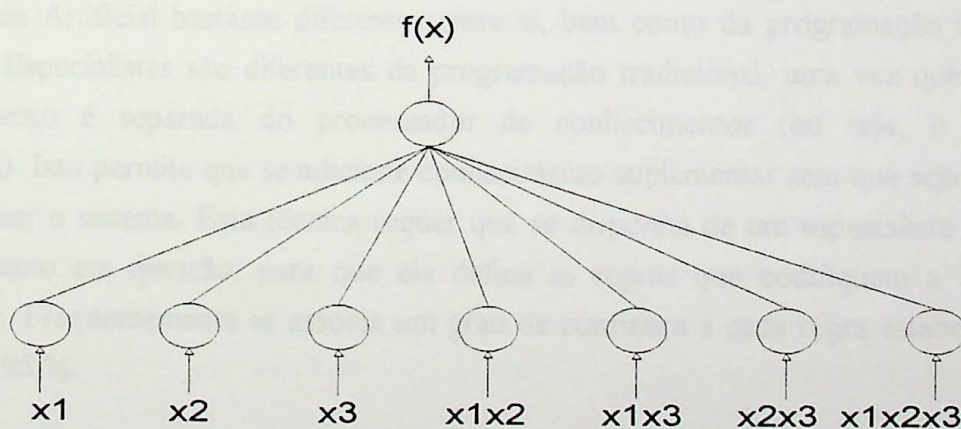
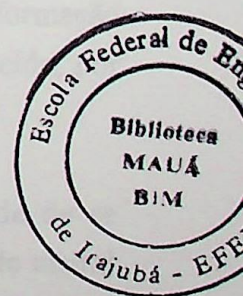


Fig. 3.10 - Modelo do produto exterior

Assim, se os valores de entrada são denotados por $\{ X_i : 1 \leq i \leq n \}$ o modelo do produto exterior poderia ter os seguintes dados de entrada: $\{ X_i, X_iX_j, X_iX_jX_k : \text{termos não duplicados} \}$.



3.6 Principais aplicações das redes neurais

Como já foi visto anteriormente, as aplicações típicas para redes neurais são aquelas onde a implementação de algoritmos ou regras de controle é impossível ou desaconselháveis. Há casos onde se desconhece um algoritmo adequado; em outros, o desenvolvimento de um programa para implementá-lo pode ser custoso e demorado. As aplicações que se enquadram nesse contexto e são particularmente adequadas para a aplicação de redes neurais são as seguintes:

- Análise e processamento de sinais
- Controle de processos
- Robótica;
- Reconhecimento de padrões
- Análise de imagens
- Análise de voz
- Diagnóstico médico
- Previsão da carga elétrica
- Processamento de alarmes
- Diagnóstico de faltas em Sistemas Elétricos de Potência
- Previsões no mercado financeiro

e muitas outras mais.

3.7 Comparação com outras técnicas de modelamento

Redes neurais, sistemas especialistas e lógica difusa ("fuzzy logic") são técnicas de Inteligência Artificial bastante diferentes entre si, bem como da programação tradicional. Sistemas Especialistas são diferentes da programação tradicional, uma vez que a base de conhecimento é separada do processador de conhecimentos (ou seja, o motor de inferência). Isto permite que se adicione conhecimento suplementar sem que seja necessário reprogramar o sistema. Esta técnica requer que se disponha de um especialista na área de conhecimento em questão, para que ele defina as regras que codifiquem a informação pertinente. Frequentemente se associa um grau de confiança a cada regra estabelecida, por exemplo: 95 %.

Já as redes neurais podem ser implementadas sem que haja a necessidade de se impor regras explícitas ou conhecimento formal prévio. Muito pelo contrário, a rede neural aprende as regras necessárias para lidar com o conhecimento, através do ajuste dos pesos relativos às sinapses, de modo a minimizar o erro entre os dados reais e os calculados pela rede.

Pode-se encarar a rede neural como um modelo muito generalizado cuja parametrização se dá através do ajuste dos pesos das sinapses. Logo, o desenvolvimento de uma rede neural prescinde do trabalho intensivo de um especialista na matéria em questão, embora ele seja necessário para definir os critérios de seleção e preparação dos dados a serem fornecidos ao treinamento da rede neural.

Além disso, há uma diferença marcante entre as redes neurais e os sistemas especialistas: nem sempre é possível saber como uma rede neural chegou a um dado resultado, enquanto que um sistema especialista sempre tem condições de dar essa informação.

Em certos casos torna-se interessante desenvolver sistemas híbridos de inteligência artificial que empregam redes neurais juntamente com sistemas especialistas ou de lógica difusa [25]. Comparando-se as redes neurais com técnicas estatísticas como, por exemplo, regressão fica evidente uma vantagem da rede neural, ou seja: não há, a princípio, a necessidade de se determinar a priori quais variáveis são importantes. As redes neurais tendem a determinar automaticamente quais são as importantes. Variáveis irrelevantes são anuladas através das reduções das energias de suas conexões com os demais neurônios a que estão ligados. Ou seja, os valores correspondentes dos pesos das sinapses são muito reduzidos. Além disso, não é necessário impor uma função como modelo, condição imprescindível para o uso da regressão estatística.

Outra vantagem das redes neurais em relação à regressão estatística é sua robustez, ou seja, maior grau de imunidade a ruídos nos dados. Aliás, é até interessante treinar as redes neurais com dados contendo ruído, para que sua precisão aumente quando em uso sob condições reais.

As características acima permitem que as redes neurais incluam o maior número possível de variáveis de entrada no modelo. Do mesmo modo que variáveis irrelevantes são progressivamente eliminadas, são criadas interações múltiplas mais complexas e sutis entre as variáveis importantes, levando à maximização da precisão no modelo. Isto é particularmente útil em modelos com mais de três variáveis de entrada envolvidas [29].

Por outro lado, o uso de técnicas estatísticas pode ser muito útil na análise preliminar dos dados, refinando a informação a ser fornecida à rede neural e, desse modo, promover a minimização do tempo e esforço requeridos em seu desenvolvimento, aumentando a precisão do modelo final.

3.8 Dimensionamento da rede neural

Ao se conceber uma rede neural, logicamente há dois parâmetros a serem definidos: o número de camadas de neurônios que ela deverá ter, bem como o número de camadas de neurônios em cada camada. Uma rede neural unidirecional deve conter no mínimo duas camadas de neurônios. Isto é óbvio: uma camada se destina à entrada dos dados e a outra à saída dos resultados. Contudo, como já pode ser visto no tópico que tratou da evolução das redes neurais, perceptrons com apenas duas camadas apresentam utilidade muito limitada. De fato, eles só são úteis se os casos a serem modelados estão separados de forma linear. Redes neurais com uma ou duas camadas ocultas, melhora o desempenho das mesmas, ou seja sua capacidade de aprendizado aumenta.

Embora a necessidade de que haja pelo menos uma camada oculta na rede seja praticamente um ponto pacífico, há considerável controvérsia quanto ao número necessário de camadas ocultas necessárias para que a rede adquira o conhecimento de forma precisa. Uma vez definido o número de camadas da rede neural, o próximo passo a ser cumprido é a definição do número de neurônios por camada.

Na camada de entrada deve haver um número de neurônios igual ao número de variáveis a serem fornecidos à rede. Normalmente se inclui à camada de entrada um neurônio com valor constante unitário, denominado "bias", que atua analogamente a um "terra" elétrico. Seu papel é aumentar o número de graus de liberdade disponíveis no modelo, permitindo que a rede neural tenha maior capacidade de se ajustar ao conhecimento à ela fornecido. A camada de saída deve conter um número de neurônios igual ao de variáveis que se deseja calcular.

O maior desafio no dimensionamento de uma rede neural é a escolha correta do número de neurônios das camadas ocultas. Na realidade, a camada oculta é uma camada generalizadora. Ela tende a combinar os neurônios das camadas de entrada e saída em meta grupos. Por exemplo, em aplicações que requeiram o reconhecimento de caracteres, a camada interna deve "aprender" a agrupar os pixels provenientes da camada de entrada em forma de linhas.

Embora a escolha correta do número de neurônios que deve conter as camadas ocultas ainda seja muito controverso, em linhas gerais, pode-se afirmar que há um número ótimo de camadas ocultas e de neurônios nelas contidas que, para cada caso, leva à maximização da capacidade preditiva da rede neural utilizada. O problema é determiná-los

com base nas poucas diretrizes hoje existentes. Isso normalmente exige um número muito grande de tentativas, se o problema for relativamente complexo [29].

3.9 Condicionamento dos dados e influência do tamanho do conjunto de treinamento

Para um treinamento bem sucedido o tamanho do conjunto de dados desempenha um papel muito importante. Devido a que a rede neural se baseia nos dados a ela fornecidas para extrair o modelo geral que rege seu inter-relacionamento, é importante que a informação a ela fornecida durante a sua fase de treinamento seja fidedigna. Caso contrário, acabará por estabelecer um modelo espúrio, pois ela não dispõe de nenhuma regra preestabelecida para poder efetuar alguma crítica aos dados a ela fornecidas.

O ideal é que a escolha dos dados para esse subconjunto se faça de forma totalmente aleatória, de modo a que ele realmente seja representativo em relação ao universo global que está sendo considerado. Um aspeto importante é apresentar à rede, durante seu treinamento, proporções iguais das diversas situações envolvidas, para que ela possa aprender como diferenciá-las de forma eficaz. Caso contrário, ela tenderá a esquecer as situações que aparecem com menor frequência.

Deve se separar de 50% a 90% do número total de registros de dados disponíveis para se obter o subconjunto que será utilizado como treinamento. O restante do conjunto de dados, 10% a 50% do total forma o subconjunto de teste. Ele é apresentado à rede neural somente após seu treinamento. Uma vez que ele foi escolhido de forma aleatória, esses dados tem grande possibilidade de serem situações diferentes das apresentadas à rede durante seu treinamento. Logo, se efetivamente a rede neural conseguiu "deduzir" corretamente o inter-relacionamento entre os dados de entrada e saída, a diferença entre os valores calculados por ela e os respectivos dados reais de saída deverá ser mínima quanto ela processar o subconjunto de teste [29].

Dadas as necessidades de subdivisão do conjunto global de dados e da máxima generalização do treinamento da rede neural, o ideal é se dispor de no mínimo algumas centenas de dados relativos à aplicação que se quer modelar.

Ao menos redes neurais somente podem processar dados numéricos. Isso não impede que outros tipos de variáveis possam ser modelados pelas redes neurais, basta apenas que eles sejam codificados previamente em forma numérica. Por exemplo: variáveis lógicas-Falso/Verdadeiro-podem ser traduzidas como 0 e 1, respectivamente.

A princípio, a magnitude dos dados não deveria afetar o desempenho das redes neurais, não é o que se verifica na prática. Contudo, tal fato poderia até ser esperado, uma vez que neurônios biológicos possuem uma faixa dinâmica bastante limitada, ou seja, processam níveis de sinais numa faixa de magnitude muito restrita.

Há ainda outra limitação em relação à magnitude dos dados. Eles devem ser adequadamente escalados para poderem ser processados nas redes neurais. De fato, perceptrons que utilizam a função sigmoideal como função de transferência trabalham numa faixa de valores entre 0 e +1, enquanto que os que operam com a função tangente hiperbólica estendem essa faixa para -1 a +1. Se os dados a serem modelados ultrapassarem a respectiva faixa, é necessário aplicar um fator de correção a eles para que sua magnitude se adeque às condições de sua rede neural, é dizer, os dados deveram ser normalizados antes de ser processados pela rede.

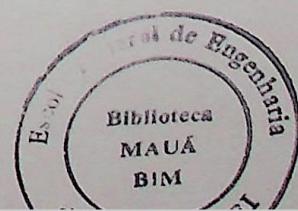
3.10 Problemas mais comuns envolvidas durante o processo de treinamento

Talvez a falha mais freqüente que ocorra durante a implementação de redes neurais seja seu mal desempenho durante sua avaliação com o subconjunto de teste, apesar do processo de treinamento ter sido aparentemente bem sucedido. O que ocorre neste caso é que a rede neural memorizou os dados fornecidos ao invés de estabelecer relações gerais entre eles. Assim ao se apresentar à rede assim treinada dados novos, ela simplesmente não consegue calcular as respostas certas. As prováveis causas deste problema são:

- Excesso de treinamento.
- Subconjunto de treinamento mal dimensionado, Rede neural superdimensionada.
- Incompatibilidade entre os subconjuntos utilizados nas fases de treinamento e teste.

As contramedidas para esses problemas são simples, mas eventualmente muito trabalhosas devido à falta de embasamento teórico para se estabelecer diretrizes seguras:

- Assegurar a qualidade dos dados fornecidos
- Repartir adequadamente os dados entre os subconjunto de treinamento e teste
- Garantir que haja no subconjunto de dados para treinamento uma partição equitativa de todos os casos que ocorrem no universo a ser modelado.
- Dimensionar criteriosamente a rede neural, efetuando-se testes com várias arquiteturas.
- Evitar excesso de treinamento, inter rompendo a intervalos regulares (a cada 100 iterações) e avaliando-se o grau de ajuste conseguido até então, utilizando-se o



subconjunto de teste. Desse modo, é possível determinar a partir de que ponto a rede passou a ficar demasiadamente treinada.

Há casos em que a rede neural falha já na fase de treinamento, ou seja, ela não converge para um estado estável e eficaz do ponto de vista preditivo. Neste caso as prováveis causas podem ser:

- Rede neural subdimensionada.
- Coeficiente de aprendizado (α) inicial muito alto.
- Dados com magnitude incompatível com a função de ativação dos neurônios.

As possíveis soluções sugeridas neste caso são:

- Dimensionamento correto da rede neural, procurando-se testar diversas arquiteturas.
- Redução na magnitude do coeficiente de aprendizado (α), promovendo-se seu declínio ao longo da fase de treinamento, de modo a evitar que os valores dos pesos das sinapses atinjam valor máximo e trunquem a capacidade de aprendizado da rede.
- Se necessário, comprimir a faixa de magnitude dos dados, e garantir que seus valores sejam adequados ao processamento pela rede neural.

3.11 Treinamento da rede neural (Back-propagation)

Vários algoritmos tem sido criados para o treinamento das redes uni direcionais (feed-forward), mas a maioria deles são versões modificadas do algoritmo do "back-propagation". Tipicamente a rede possui uma camada de entrada, uma camada de saída, e uma ou várias camadas ocultas, como pode ser visto na fig 3.11. Não existe limite teórico enquanto ao números de camadas ocultas, porém tipicamente poderiam ser usados uma ou duas camadas.

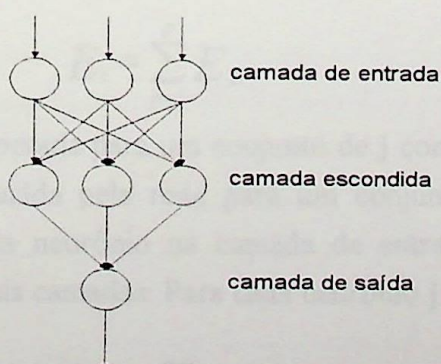


Fig. 3.11 - Rede unidirecional

O treinamento da rede neural uni direcional envolve a otimização dos pesos das sinapses, ou seja, as energias correspondentes às ligações entre os neurônios, de modo a fazer com que ela mapeie corretamente a função proposta entre as variáveis de entrada e de saída. O aprendizado deste tipo de rede se faz pelo método de retro propagação ("back-propagation"). O processo é iterativo. A cada iteração durante a fase de treinamento, é feita uma comparação entre os valores reais de saída do subconjunto de dados de treinamento com os correspondentes valores calculados pela rede neural a partir dos dados de entrada.

Em função da magnitude das diferenças (erros) assim constatadas, o algoritmo de aprendizagem recalcula os pesos das sinapses de modo a minimizar os desvios. O erro na camada de saída é propagado para atrás até a camada de entrada.

3.12 Algoritmo Back-propagation

O algoritmo de aprendizado "Back-propagation" é o método mais frequentemente utilizado no treinamento das redes neurais, e é proposta neste trabalho como uma metodologia na (PCCP) previsão da carga à curto-prazo. O (GDR) "Regra Delta Generalizada" [36] sugerido por Rumelhart é utilizado para treinar a rede neural tipo perceptron. Neste tipo de rede um vetor de saída é produzido pela apresentação de um modelo de entrada na rede. Através das diferenças entre as saídas esperadas e produzidas, os pesos da rede $\{W_{ij}\}$ são ajustados para reduzir o erro na saída. O erro na camada de saída é propagado para atrás até alcançar as camadas de entrada. Devido à retro propagação de erro, o GDR também é chamado de algoritmo back-propagation.

Especificamente, o erro medido sobre o modelo \mathbf{p} temos:

$$E_p = \frac{1}{2} \sum_{j=1}^M (t_{pj} - o_{pj})^2 \quad (1)$$

e o erro total de todas as medidas como:

$$E_t = \sum_{p=1}^p E_p$$

onde t_{pj} é a saída esperada para um conjunto de \mathbf{j} componentes de saída para o modelo \mathbf{p} e O_{pj} é a saída produzida pela rede para um conjunto de \mathbf{j} componentes do modelo de entrada \mathbf{p} . Para cada neurônio na camada de entrada, o neurônio de saída atua como entrada para as demais camadas. Para cada neurônio \mathbf{j} na camada oculta, a entrada da rede é dado por :

$$net_j = \sum_i w_{ji} o_i \quad (2)$$

Onde i é o neurônio na camada precedente, O_i é a saída do nodo i , e W_{ji} é a conexão dos pesos do neurônio i para o neurônio j (figura 3.12).

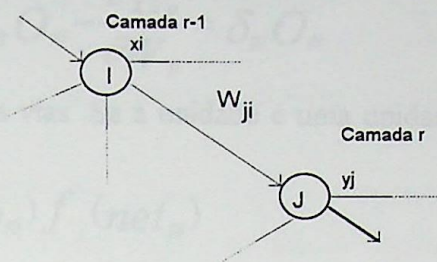


fig. 3.12 - conexão entre dois neurônios

O neurônio da saída é dado por :

$$O_j = \frac{1}{1 + \exp - (net_j + \theta_j)} = f_j(net_j, \theta_j) \quad (3)$$

onde net_j é o sinal de entrada para o nodo j e θ_j é o bias. No processo de treinamento, θ_j pode ser considerado como uma conexão entre o nodo j e um nodo fictício de saída (bias) permanecendo sempre como unitário. A figura 3.13 ilustra o modelo matemático do neurônio

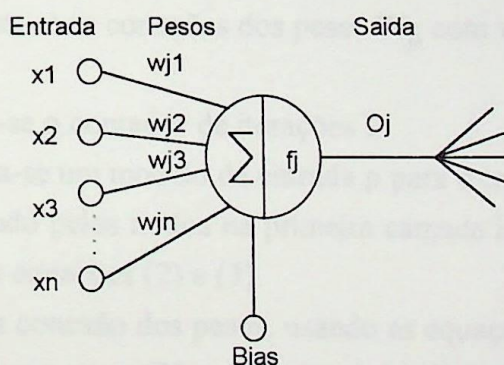


Fig. 3.13 - modelo matemático do neurônio

f_j é uma função diferencial e não decrescente e W_{ji} é o peso a ser ajustado. A função f_j é normalmente uma função sigmoidal como mostra a figura 3.14

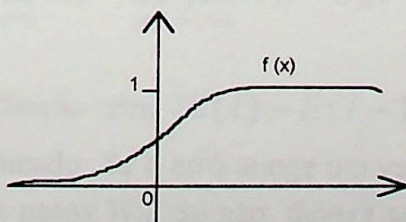


Fig. 3.14 - função sigmoidal

Para se obter uma regra de ajuste dos pesos, o gradiente de E_p em relação a W_{ji} é utilizado e representado como segue:

$$-\frac{\partial E_p}{\partial W_{ji}} = \delta_{pj} O_{pi} - \frac{\partial E_p}{\partial W_{ji}} = \delta_{pj} O_{pi} \quad (4)$$

onde δ_{pj} é definido em duas vias. Se a unidade é uma unidade de saída, temos a seguinte expressão:

$$\delta_{pj} = (t_{pj} - o_{pj}) f'_j(\text{net}_{pj}) \quad (5)$$

e para uma unidade numa determinada camada interna temos:

$$\delta_{pj} = f'_j(\text{net}_{pj}) \sum_k \delta_{pk} w_{kj} \quad (6)$$

onde f'_j é a derivada de f_j . A regra de ajuste dos pesos pode ser obtido usando a equação (4) onde temos:

$$\Delta W_{ji}(n+1) = \eta \delta_{pj} o_{pi} + \alpha \Delta W_{ji}(\eta) \quad (7)$$

onde η é o parâmetro de aprendizado e α é o momentum constante para determinar os efeitos das mudanças de pesos anteriores. Uma vez que a rede neural foi treinada ela produz saídas muito rápidas para um conjunto de dados de entrada. O algoritmo é descrito pelos seguintes passos.

- Passo 1: Inicializam-se as conexões dos pesos W_{ji} com valores aleatórios pequenos entre -0.5 e +0.5
- Passo 2: Inicializa-se o contador de iterações I.
- Passo 3: Apresenta-se um modelo de entrada p para o treinamento.
- Passo 4: Começando pelos nodos na primeira camada interna, calcular o neurônio de saída usando as equações (2) e (3).
- Passo 5: Adaptar a conexão dos pesos, usando as equações (5), (6), e (7).
- Passo 6: Repetir os passos (3) a (5) até que todos os modelos no treinamento tenham sido considerados ($p = P$).
- Passo 7: Calcular o erro $E(I)$ para a presente iteração I usando a seguinte equação:

$$E(I) = \sum_{p=1}^P E_p = \sum_{p=1}^P \frac{1}{2} \sum_{j=1}^M (t_{pj} - o_{pj})^2$$

- Passo 8: Calcular a função erro $|E(I) - E(I-1)| < \varepsilon$, onde ε é um número pequeno positivo e estimado. Se o erro atinge um valor estimado, o processo para e é imprimido fora os pesos W_{ji} . Se não, devera ser incrementado o número de iterações I e repetir o passo 3.

Através da variação do erro médio total pode-se ter uma idéia do comportamento do treinamento, ou seja se a rede esta armazenando conhecimento e uma noção sobre a velocidade do mesmo. O acompanhamento do erro total pode, portanto, evitar perda de tempo com treinamentos inadequados e mal projetados.

3.13 Algoritmo do GMDH "Group Method of Data Handling"

O algoritmo do GMDH foi primeiramente proposto por Ivakhnenko como um método geral, e conveniente para previsão e estimação de parâmetros. Desde então foram publicadas várias literaturas sobre esta aplicação para problemas práticos nas áreas de reconhecimento, identificação, e estimação de parâmetros [37]. O algoritmo permite modelar sistemas complexos sem ter conhecimentos específicos do sistema ou de uma quantidade massiva de dados. O algoritmo funciona da seguinte maneira:

Sendo X_1, X_2, \dots, X_n , as variáveis de entrada do sistema, elas são combinadas para ter a próxima geração de variáveis. Começamos calculando a equação de regressão Y para cada par de variáveis de entrada X_i e X_j através da fórmula:

$$y = A + Bx_i + Cx_j + Dx_i^2 + Ex_j^2 + Fx_ix_j$$

ver a fig. 3.15

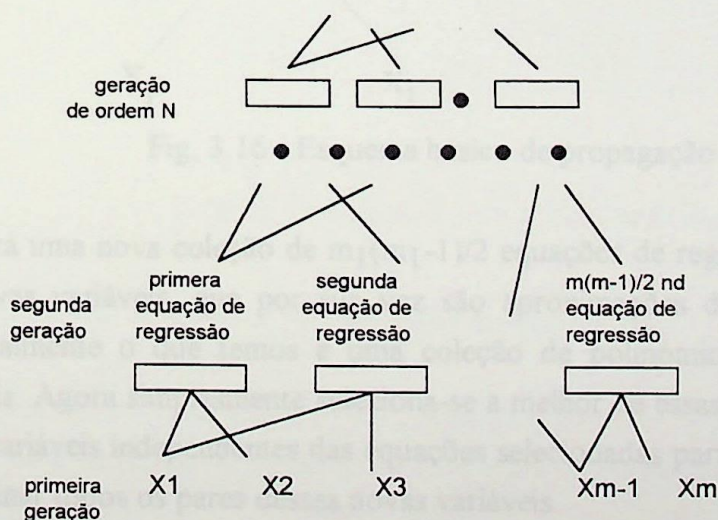


Figura 3.15 - formação da geração de variáveis tomadas duas por vez

isto vai dar um total de $m(m-1)/2$ combinações de variáveis para a previsão na saída Y em lugar das m variáveis originais X_1, X_2, \dots, X_n . Após encontrar essas equações de regressão

de um conjunto de observações de entrada/saída, procuramos quais variáveis devem ser salvas. Isto vai dar uma coleção de modelos de regressão quadráticos (m_1) que melhor aproxima à saída Y (notar que cada previsão depende de duas variáveis independentes).

Agora utiliza-se cada uma dessas equações quadráticas para calcular e gerar novas variáveis independentes, que irão substituir às variáveis originais (X_1, X_2, \dots, X_n). Essas novas variáveis independentes poderiam ser combinadas e gerar outras variáveis e assim pela frente. Isto é, calcula-se todas as equações de regressão quadráticas de Y versus essas novas variáveis (duas por vez) ver fig. 3.16

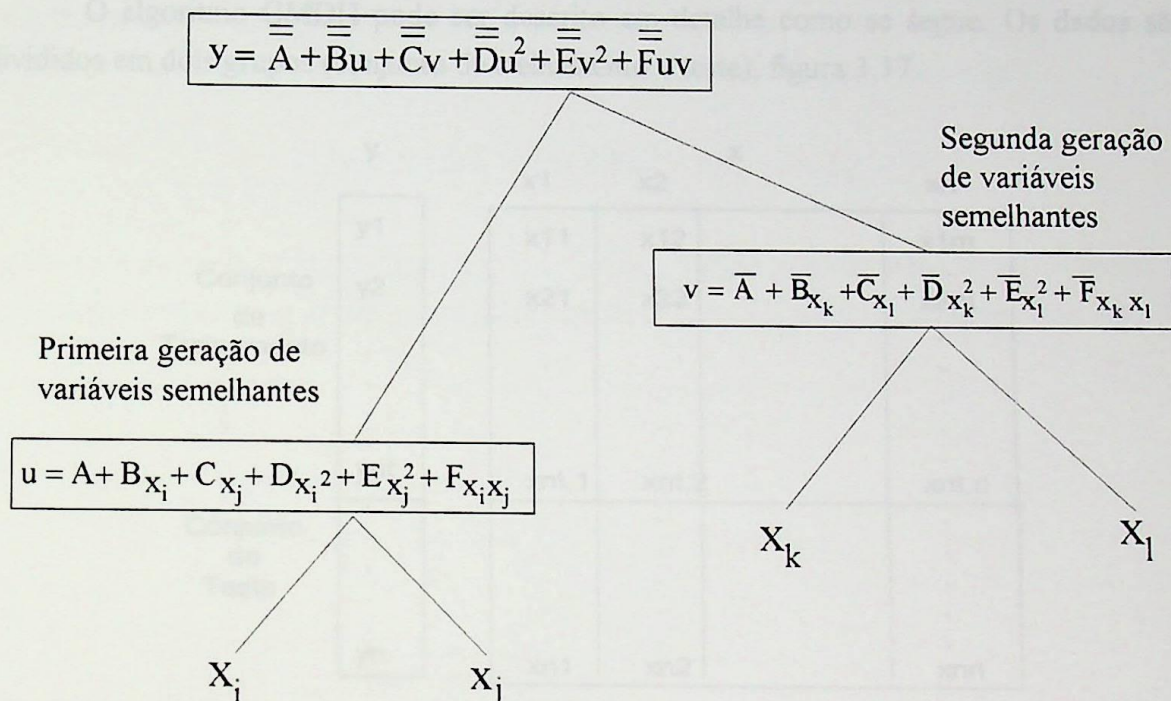


Fig. 3.16 - Esquema básico de propagação de variáveis

Isto dará uma nova coleção de $m_1(m_1-1)/2$ equações de regressão para prever a saída Y das novas variáveis, que por sua vez são aproximações de Y das equações anteriores. Essencialmente o que temos é uma coleção de polinômios de quarto grau em quatro variáveis. Agora simplesmente seleciona-se a melhor de essas novas aproximações, gerando novas variáveis independentes das equações selecionadas para substituir as variáveis velhas, e combinar todos os pares dessas novas variáveis.

O processo continua até a equação de regressão comece a ter uma predictabilidade pobre que outras equações prévias. Isto ocorrerá testando a equação de regressão versus novas observações independentes. Após parar o treinamento, é escolhido os melhores polinômios quadráticos nessa geração.

O que temos na verdade é uma aproximação da saída y como um quadrado de duas variáveis, que a sua vez são os mesmos quadrados de mais de duas variáveis, que são os quadrados das variáveis originais. Em outras palavras, se fazemos as substituições algébricas necessárias, chegaremos num complicado polinómio da forma:

$$y = a + \sum_{i=1}^m b_i x_i + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m d_{ijk} x_i x_j x_k + \dots$$

que é conhecido como o **Polinómio de Ivakhnenko**

O algoritmo GMDH pode ser descrito em detalhe como se segue: Os dados são divididos em dois grupos (conjunto de treinamento e teste), figura 3.17.

	y	x			
		x1	x2		xn
Conjunto de Treinamento	y1	x11	x12	x1m
	y2	x21	x22		x2m

	ynt	xnt,1	xnt,2		xnt,n
Conjunto de Teste

	yn	xn1	xn2		xnn

fig. 3.17 - entrada para o algoritmo do GMDH

- Passo 1 (Construção de novas variáveis Z_1, Z_2, \dots, Z_n)

O primeiro passo consiste em tomar simplesmente todas as variáveis independentes (colunas de x) duas por vez e construir a equação de regressão polinomial forma:

$$y = A + Bu + Cv + Du^2 + Ev^2 + Fuv$$

que melhor ajuste às observações dependentes y_i no treinamento. Em outras palavras, cada uma das regressões polinomiais $\binom{m}{2} = m(m-1)/2$ poderiam ser calculadas das observações

$\{(y_i, x_{ip}, x_{iq}) : i = 1, \dots, nt\}$ para diferentes $p, q = 1, \dots, m$. Essas $m(m-1)/2$ superfícies de regressão são ilustrados na figura 3.18

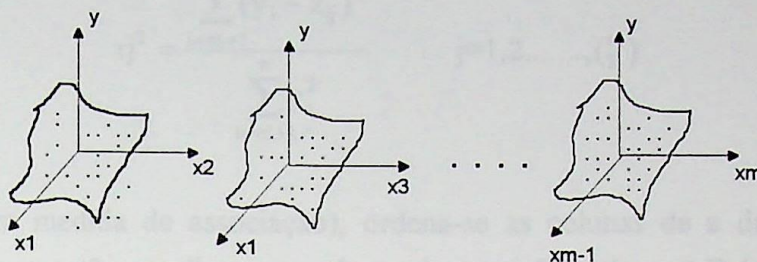


fig. 3.18 - construção de novas variáveis

Agora, para cada uma dessas superfícies de regressão, é avaliado o polinômio para todos os n pontos. Por exemplo, na primeira superfície mostrada acima, pode-se avaliar

$$y = A + Bx_1 + Cx_2 + Dx_1^2 + Ex_2^2 + Fx_1x_2$$

sobre os n dados $(x_{11}, x_{12}), (x_{21}, x_{22}), \dots, (x_{n1}, x_{n2})$ e guardar essas novas observações (novas gerações de variáveis) na primeira coluna de uma nova formação z , ver figura 3.19

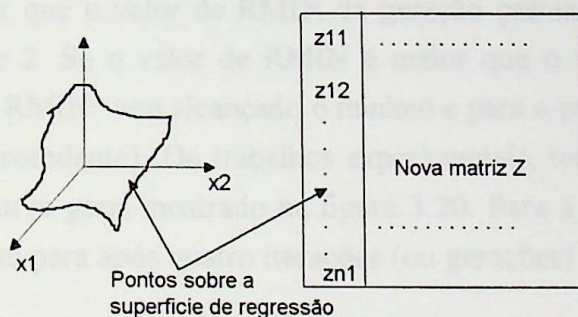


fig. 3.19 - Construção de novas variáveis

Os resultados da coluna z são calculados de uma maneira similar. Pode-se interpretar essas novas variáveis como sendo as variáveis que tem melhor previsibilidade que todas aquelas variáveis da geração original x_1, x_2, \dots, x_n . O objetivo é guardar somente a melhor de essas novas variáveis (seleção das melhores), e aqui é onde o conjunto de teste entra a desempenhar um papel importante (até aqui não foi utilizado essas variáveis).

- **Passo 2** (Investigar fora as mínimas variáveis eficazes)

Este passo substitui as colunas de x (variáveis velhas) pelas colunas de z (novas variáveis) que melhor aproxima à variável independente y no conjunto de teste. Isto é, para

cada coluna j de z é calculado a raiz do mínimo quadrado (também chamado de critério de regularidade) r_j

$$r_j^2 = \frac{\sum_{i=nt+1}^n (y_i - z_{ij})^2}{\sum_{i=nt+1}^n y_i^2} \quad j=1,2,\dots,(m)$$

(ou alguma outra medida de associação), ordena-se as colunas de z de acordo com o crescimento de r_j , e então escolhe-se as colunas de z satisfazendo $r_j < R$ (onde R é número fixo escolhido pela frente) para substituir as colunas originais de x (essas colunas de z serão a nova coluna x). Pode-se notar também que o número de variáveis salvas (m_1), poderia ser menor ou maior que o número de variáveis originais (m). Também poderia ser notado que o teste do valor apropriado r_j foi calculado sobre as observações no "conjunto de teste".

- Passo 3 (Teste de otimização)

Do passo 2 encontramos os menores valores de r_j e isto é chamado de RMIN (mínimo quadrado). Então plota-se o RMIN num gráfico como mostra a figura 3.20. Se o valor de RMIN é menor que o valor de RMIN da geração prévia, então retrocedemos e repetimos os passos 1 e 2. Se o valor de RMIN é maior que o valor precedente, então assumimos que a "curva RMIN" tem alcançado o mínimo e para o processo (e é utilizado os resultados da geração precedente). De trabalhos experimentais, tem sido verificado que a curva RMIN possui a curva geral mostrado na figura 3.20. Para a curva RMIN específico da figura 3.20, o processo para após quatro iterações (ou gerações).

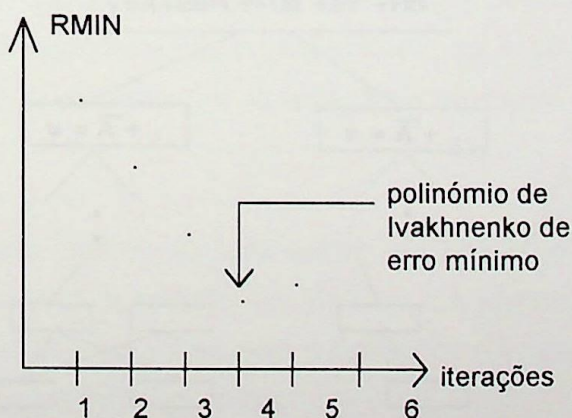


fig.3.20 - Critério de parada

A primeira coluna de z terá os valores \bar{y}_i do polinómio de Ivakhnenko

$$\begin{aligned}\bar{y}_1 &= a + \sum_{i=1}^m b_i x_{1i} + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_{1i} x_{1j} + \dots \\ \bar{y}_2 &= a + \sum_{i=1}^m b_i x_{2i} + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_{2i} x_{2j} + \dots \\ &\vdots \\ \bar{y}_n &= a + \sum_{i=1}^m b_i x_{ni} + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_{ni} x_{nj} + \dots\end{aligned}$$

avaliada sobre os n dados originais. Em outras palavras, a primeira coluna de z é o preditor da variável dependente. Para encontrar os coeficientes a , b_i , c_{ij} , d_{ijk} , no polinómio de Ivakhnenko, deveriam ser salvas todas as superfícies da regressão quadrática

$$y = A + Bu + Cv + Du^2 + Ev^2 + Fuv$$

que foram calculadas a cada iteração e sistematicamente avaliar esta árvore de segunda ordem até chegar ao polinómio de alto grau nas variáveis originais X_1, X_2, \dots, X_n . Do ponto de vista computacional, uma vez que tem-se todas as regressões polinomiais quadráticas armazenadas no computador, é possível calcular a previsão \bar{y} da saída y desses quadrados, figura 3.21.

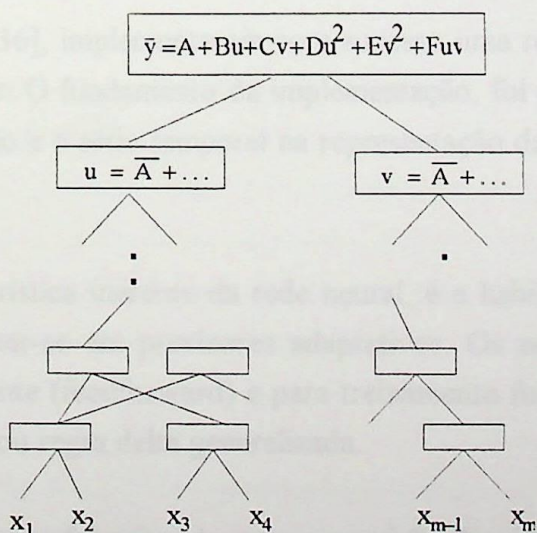


fig. 3.21 - Árvore polinomial

CAPÍTULO 4

Redes Neurais em Previsão de Carga

4.1 Introdução

Recentemente, muito interesse tem sido fixado sobre a aplicação de RNA's na previsão da carga [36]. Nesses estudos, a técnica de RNA's tem sido aplicada para correlacionar as informações climáticas e as variações da carga para prever picos de carga diária, carga global e carga horária. Os resultados neste campo são muito promissores. Diversos são os artigos publicados ultimamente, dos quais uma breve revisão bibliográfica dos principais trabalhos publicados será feita a seguir.

4.2 Revisão Bibliográfica

G. Lambert Torres *et al.* [18] implementou um modelo híbrido de rede neural e sistema especialista difuso. A rede neural era do tipo por camada e foi treinada utilizando-se a técnica de retropropagação do erro. Uma primeira previsão era feita por esta rede que, em seguida, entregava este valor ao sistema especialista difuso para um ajuste fino da previsão. Redes com diversos números de neurônios foram treinadas e os melhores resultados foram apresentados no artigo.

Park *et al.* [36], implementaram com sucesso uma rede neural artificial como parte principal do previsor. O fundamento da implementação, foi a possibilidade de combinar os métodos de regressão e a série temporal na representação de um modelo não linear de uma série de dados.

Uma característica inerente da rede neural, é a habilidade do aprendizado próprio, que faz a rede basear-se em previsores adaptativos. Os autores utilizaram o modelo de propagação para frente (feedforward) e para treinamento foi utilizado o algoritmo de retro propagação de erro ou regra delta generalizada.

A saída da rede foi utilizada para a previsão de picos diários de carga, e previsão horária de carga. Para cada aplicação foi utilizada a mesma configuração, baseado somente numa camada interna. No treinamento foram utilizados dados reais da Puget Sound Power

and Light Company, correspondentes a um período de três meses. Os erros médios encontrados foram de 2,04% e 1,68% para picos diários e carga horária, respectivamente.

O artigo publicado por **Peng et al.** [38] emprega uma simples rede neural com uma camada interna para a previsão. Eles utilizaram uma função de ativação sigmoideal para os neurônios da camada interna e uma função de ativação linear para os neurônios da camada de saída. As entradas foram representadas por cinco neurônios, representando cargas, temperaturas máximas e mínimas das horas passadas e presente, e a saída da rede representando a carga horária. Seus resultados mostraram erros médios na ordem de 13% (erros médios de 4% a 5% são aceitos na prática). Os autores chegaram à conclusão que a causa do erro maior se devia ao fato de que dias especiais (fim de semana) eram tratados como dias normais durante o treinamento.

Um outro sucesso da aplicação de rede neurais na previsão da carga foi dado por **Lee et al.** [39]. Esses autores implementaram uma rede neural com algoritmo de treinamento de retropropagação do erro para predizer a carga horária um passo à frente num período de 24 hrs. Neste sentido dois diferentes métodos foram aplicados : o primeiro, com um passo à frente da previsão, chamado de método estático, e o segundo, com 24 passos à frente, chamado de método dinâmico.

Em ambos os métodos, a carga foi dividida em duas categorias, dia de semana e fim de semana (sábado e domingo). Várias arquiteturas foram testados para os diferentes métodos. A percentagem do erro médio relativo foi de 1,89% para o primeiro e 1,83% para o segundo método, que foi comparado com 1.40% obtido pelas técnicas clássicas com os mesmos dados de entrada.

Um dos aspectos mais relevantes da seleção de dados para a rede neural tem sido considerado por **Peng et al** [40]. Neste trabalho, os autores propuseram um método (distâncias medidas) que tem sido tradicionalmente utilizado para a identificação de agrupamento de dados, para obter casos de treinamento apropriados. Utilizando este método, 6.2% de erro foi obtido, quando incluídos fins de semana no treinamento, e 5% quando foram excluídos os fim de semanas. Para o treinamento e teste foi utilizado históricos de carga de um ano.

O artigo de **Ho et al.** [41] reporta vários resultados experimentais da aplicação de rede neurais na previsão da carga para 24 hrs. Neste artigo, os autores utilizaram o algoritmo convencional de retropropagação do erro aplicado a uma rede neural com uma camada interna. Eles examinaram a influência de vários fatores a fim de obter melhor

convergência do processo de aprendizado e melhorar a precisão da previsão. Os resultados revelaram que a convergência do algoritmo convencional de retropropagação do erro é fortemente afetada pelas taxas de momento e de treinamento. Também examinaram os efeitos da apresentação dos dados de treinamento e vários conceitos na convergência do algoritmo de aprendizado.

Como resultado da experiência, os autores propuseram um novo algoritmo de aprendizado que atualiza o momento e a taxa de aprendizado durante o aprendizado. O algoritmo proposto foi testado com os dados da Taiwan Power Company e os resultados previstos foram comparados com previsões feitas pelos operadores e baseados em sistemas especialistas. Concluindo que as previsões geradas pelo novo algoritmo foram melhores que as previsões geradas pelos operadores e baseados em sistemas especialistas, tendo um erro médio absoluto de 0,7% a 0,9%. Um outro artigo que podemos destacar nesta linha é o de **A.P. Alves da Silva *et al***, [21].

O artigo do **Shin-Tzo Chen *et al*** [42], apresenta um modelo de rede neural para prever carga sensível às mudanças climáticas. O modelo proposto é capaz de prever uma semana inteira. O modelo não está totalmente conectado, possuindo menor tempo de treinamento do que a rede neural completamente conectada. O referido modelo proposto pode diferenciar cargas entre fim de semana e dia de semana. Os resultados mostram que o modelo pode permitir previsões com maior precisão do que o modelo estatístico convencional.

No artigo do **Yuan-Yih Hsu *et al*** [43,44] um novo método usando redes neurais é apresentado para previsão de carga. Na primeira parte do artigo uma rede neural baseada em mapas característicos de organização própria é utilizada para identificar os dias com modelos de curvas de carga similares. Na segunda parte do artigo uma rede do tipo multi camadas com aprendizado supervisionado é desenvolvida para prever picos máximos e mínimos de curva de carga. Os resultados de previsão de carga a curto-prazo da Taiwan Power System mostraram o bom desempenho da rede neural proposta.

No artigo de **Alex D. Papalexopoulos *et al*** [45], apresenta o desenvolvimento e a implementação de um modelo de Rede Neural para previsão de carga a curto prazo para o Energy Control Center da Pacific Gas & Electric Company (PG&E). Muitas ações foram tomadas para modelar com precisão eventos especiais, tais como feriados, dias quentes, frente fria e outros distúrbios do modelo normal da carga. O modelo tem sido testado sob uma variedade de condições mostrando excelentes resultados.



No artigo de **Alvaro García Tejedor *et al*** [46], implementaram uma Rede Neural para previsão horária de carga através de dois passos: classificação de dias típicos através de mapas característicos de organização própria e a previsão usando uma rede neural multi camadas. Os resultados mostram boa precisão considerando classificação de dias típicos.

No artigo de **T.S.Dillon *et al*** [7], tem sido desenvolvido dois novos métodos de previsão de carga que utiliza métodos de organização própria e técnicas de reconhecimento de padrões. A aplicação desses dois métodos para um problema prático mostra que são viáveis computacionalmente e oferecem bons resultados.

4.3 Modelo proposto

A proposta deste trabalho é fazer um estudo comparativo de previsão de carga, utilizando as técnicas das Redes Neurais Artificiais e métodos estatísticos clássicos (Box & Jenkins). Também serão testados dois algoritmos de treinamento utilizados pela rede neural, a retropropagação do erro e o GMDH (Group Method of Date Handling). Um outro estudo de interesse neste trabalho é comparar os resultados da previsão considerando cargas globais e por agrupamento. Para tal efeito foram utilizados dados de carga da Hydro-Quebec do Canadá.

Primeiramente, um algoritmo de agrupamento é utilizado para a classificação dos dias típicos. Posteriormente, uma rede neural do tipo multi-camadas (feedforward) é utilizada para avaliar a previsão. A seguir, considerações sobre classificação de dias típicos são abordadas.

Para obter um modelo de carga próprio para um certo dia, históricos de dados de carga são analisados para dividir todos os dias dentro de vários grupos, com cada grupo formado pelos dias com modelos de carga similares. Esses dias com modelos de carga similares são referidos como do mesmo dia típico. Uma vez que todos os dias típicos têm sido identificados dos dados passados de carga, cada dia incluindo o dia em estudo pode ser associada um dia típico.

Situações anômalas (mudanças de temperatura, feriados, etc) não podem ser eficientemente tratadas devido a que não estão correlacionadas com dias típicos. Tratamento de situações anômalas implica em identificação antecipada para que um procedimento de previsão específico possa ser desenvolvido.

Assim a idéia da identificação de dias típicos surgem como uma fase prévia para a própria previsão. Toda a informação de um certo dia (relação com dias precedentes, influências das variáveis climáticas) está contida na curva de carga. Um algoritmo de agrupamento baseado em distâncias euclidianas foi desenvolvido para a identificação dos dias típicos. Neste algoritmo cada agrupamento é definido por um protótipo p , que pode combinar características de mais de um exemplo de treinamento (modelo de carga). O algoritmo não especifica o número de agrupamentos; este usa um parâmetro de vigilância que controla a granularidade dos grupos. Este tipo de algoritmo é classificado como um algoritmo não-supervisionado. O algoritmo é descrito a seguir:

Passo 1: Início do processo de treinamento: nenhum protótipo foi gerado, ou seja, não existem unidades ativadas na camada de saída.

Passo 2: Repetir até que nenhum exemplar de treinamento (padrão de entrada) cause mudança no conjunto dos vetores protótipo. Neste ponto finalize o algoritmo porque estabilidade foi alcançada. Escolha um exemplar de treinamento x^k (em ciclos).

Passo 3: Encontre o protótipo P_i (se algum) mais próximo (distância Euclidiana) do exemplar de treinamento x^k .

Passo 4: Teste se o protótipo P_i é suficientemente similar ao exemplar x^k .

4a. Se for então

- o exemplar x^k pertence ao grupo representado pelo protótipo P_i .
- modifique P_i para ser mais parecido com o exemplar x^k .
- volte ao passo 2 para escolher um novo exemplar.

4b. Caso contrário

- o exemplar P_i não pertence a nenhum grupo existente. Faça um novo grupo e o protótipo deste grupo será o exemplar de treinamento x^k .
- retorne ao passo 2 para escolher um novo exemplar.

Após a classificação de dias típicos, o próximo passo é avaliar a previsão de carga usando uma rede neural. A estrutura usada pela rede neural consiste de três camadas (camada de entrada, camada escondida, e camada de saída). A camada de entrada é composta por variáveis de carga. Outras variáveis como temperatura e velocidade do vento podem ser incluídas.

A camada de saída é composta por um neurônio. Esse neurônio fornece o valor da carga prevista. Vale enfatizar que o horizonte de previsão é feito um passo à frente. O número de neurônio da camada escondida é empiricamente uma média entre o número de neurônios das camadas de entrada e de saída. Os pesos são calculados por um processo de aprendizado usando o algoritmo de retropropagação do erro. O aprendizado é feito usando medidas de dados de carga num barramento. Uma entrada inicial de dados é apresentada para a rede neural que ajusta o valor dos pesos para um erro mínimo. Em seguida, uma nova entrada de dados é apresentada e os valores dos pesos são ajustados novamente. O processo é interrompido quando a diferença entre o valor de saída real e previsto para todo o conjunto é menor que um valor pré-estipulado.

4.3.1 *Influência do Agrupamento*

O modelo da carga é um reflexo das atividades da população. Assim tem-se características bem definidas para certos dias e certas épocas do ano. Modelos de curvas de cargas de fim de semana e dias de semana apresentam características bem diferentes. Também modelos para verão e inverno possuem comportamentos diferentes, por exemplo, no Canadá, onde tem-se grandes variações de temperatura.

Fazendo-se a previsão considerando as cargas não agrupadas (um único conjunto de dados), a margem de erro geralmente é muito elevado, devido às características dos diferentes tipos de dias que são tratados. Esses fatores tornam o modelo de previsão muito complexo. Neste caso o agrupamento tem um papel muito importante na previsão da carga. Agrupando-se modelos de curvas similares tem-se modelos bem mais definidos.

4.3.2 *Influência do Destendenciamento*

Um aspecto vital de processamento de séries temporais, especialmente para os métodos estatísticos, é que os componentes determinísticos devem ser eliminados. Isto inclui variações sazonais e tendências. Este trabalho mostra que a eliminação da tendência e da sazonalidade também é importante para uma boa previsão com redes neurais. As razões por que se analisa esta componente é que permite elaborar modelos descritivos do passado e uma vez identificada a tendência ela pode ser eliminada para que se possam isolar as componentes sazonais e cíclicas. Existem métodos de cálculo da tendência: métodos gráficos, simples mas pouco precisas; e métodos analíticos, com maior grau de precisão por darem à tendência uma expressão analítica a partir do qual se podem fazer previsões. A tendência pode ser linear ou não linear, crescente, decrescente ou constante.

4.3.3 Influência da sazonalização

A dependência sazonal (sazonalidade) é outro componente geral do modelo das séries temporais. Isto é definido como a dependência correlacional de ordem k entre cada elemento da série (correlação entre dois termos); onde k é usualmente chamado de atraso (lag). Se o erro medido não é tão grande, a sazonalidade pode ser visualmente identificado na série como um modelo que se repete a cada k elementos. A dependência entre os valores da série para um atraso particular k pode ser removido pela diferenciação das séries. A razão de tal transformação é que pode ser identificado a natureza interna da dependência sazonal das séries. A outra razão para remover a dependência sazonal é fazer a série "estacionária" que é necessário para o modelo ARIMA e outras técnicas.

4.3.4 Influência da transformação Box-Cox

A transformação Box-Cox transforma dados de séries temporais de modo a estabilizar a variância das séries, a fim de tornar uma série estacionária. Para os valores originais da série Z , a transformação ZT é dada por [47]:

$$ZT = \frac{(Z + \lambda_2)^{\lambda_1} - 1}{\lambda_1 g(\lambda_2 - 1)} \quad \text{se } \lambda_1 > 0$$

$$ZT = g \ln(z + \lambda_2) \quad \text{se } \lambda_1 = 0$$

onde g é a média geométrica da amostra de $Z + \lambda_1$. O primeiro parâmetro λ_1 governa a solidez da transformação. $\lambda_1 = 1$ corresponde aos dados originais, e $\lambda_1 = 0$ a um logaritmo. λ_2 é uma simples adição aos dados após λ_1 ser aplicado. Os valores são escalados pela média geométrica para ter a variância constante.

4.3.5 Modelo predictor de Redes Neurais (Feed-Forward)

Alguns tipos de redes neurais são capazes de aceitar vetores de valores reais em suas dimensões originais como entrada e produzir saídas de valores reais, para ser utilizados em previsão com séries temporais, como no caso do GMDH (Group Method of Data Handling). Outras redes precisam que os valores reais sejam escalados primeiramente antes de serem utilizados pela rede neural como entrada, como por exemplo a retropropagação do erro que pode utilizar na saída a função sigmoïdal cujo faixa dinâmica muito limitada encontram-se entre os valores de 0 e 1. Aqui não será tratado um modelo de rede em particular. Será assumido uma estrutura simples possível. A rede neural é uma caixa preta, que aparte de definir a arquitetura geral da rede, o usuário não tem uma visão sobre o

progresso do treinamento [52]. Ela está composta de uma ou mais entradas e uma saída que fornece o valor previsto um passo a frente, uma estrutura típica de uma rede neural para previsão de carga com séries temporais é resumido na figura 4.1. A figura mostra o uso de vários pontos contíguos ($p(k)$, $p(k-1)$, $p(k-2)$, $p(k-3)$, $p(k-4)$, $p(k-5)$,....., $p(k-n)$) de uma série temporal para prever o próximo ponto $p(k+1)$, um passo a frente. Uma série de treinamentos são utilizados para gerar um grande número de exemplos individuais. Cada exemplo do modelo mostrado poderia consistir de oito (8) pontos, por exemplo: o próximo ponto $p(k-n)$, seis pontos de histórico de carga ($p(k)$,....., $p(k-5)$), e a saída ($p(k+1)$), que poderia ser usado para fornecer o valor previsto. Usualmente vários passos de oito (8) pontos contíguos disponíveis no treinamento das séries poderiam ser utilizados para treinamento. Como pode ser visto o treinamento poderia começar a ser muito grande. É também possível fazer uso de mas de uma série de treinamento para gerar os passos de treinamento. Se o uso de vários grupos contíguos de pontos resulta num excessivo número de exemplos de treinamento, é razoável utilizar subpasso. Um melhor método e rápido é seleccionar exemplos de treinamento aleatoriamente do conjunto das séries.

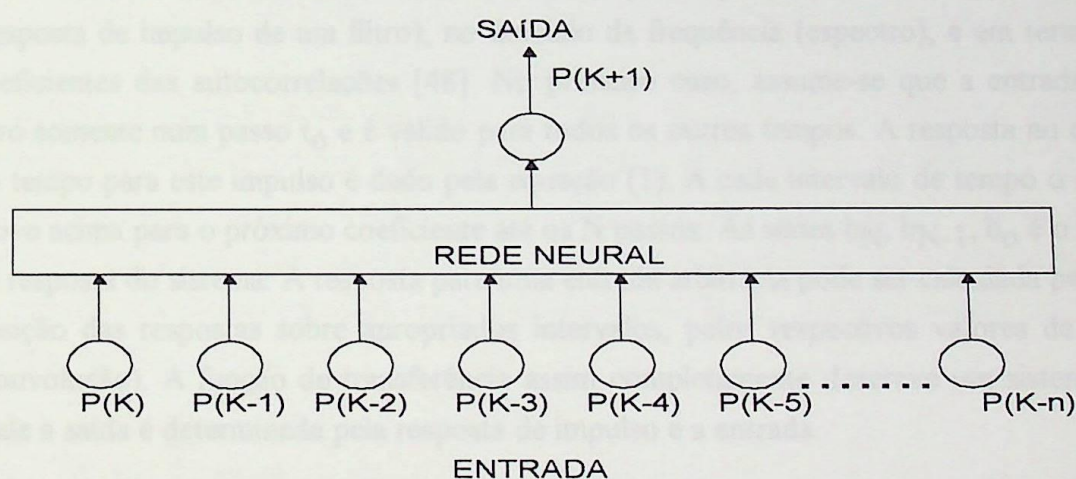
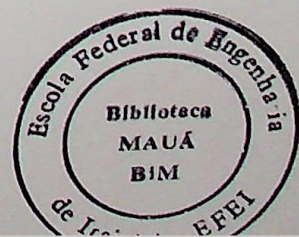


Figura 4.1 - Modelo de rede neural para previsão de carga com séries temporais.

4.3.6 Modelo linear de séries temporais

Os modelos lineares de séries temporais têm duas características: eles são entendidos em detalhe e facilmente implementados. A dificuldade é que eles podem ser inteiramente inapropriados para modelar sistemas complexos. Nesta seção será feita uma revisão das características básicas de tais modelos e explicado porque os modelos lineares podem falhar.



4.3.7 Propriedades de um modelo linear

A série será assumida como sendo linear e estacionária

Modelo média móvel (MA) Dada uma série temporal $\{e_t\}$ e desejando modificar esta série para produzir outra série $\{x_t\}$ o presente valor de x é influenciado pelo presente e passado dos N valores da série de entrada e . A relação entre a entrada e saída é dada por:

$$x_t = \sum_{n=0}^N b_n e_{t-n} = b_0 e_t + b_1 e_{t-1} + \dots + b_N e_{t-N} \quad (1)$$

Esta equação descreve o filtro de convolução: a nova série x_t é gerada por um filtro de ordem N com coeficientes b_0, \dots, b_n da série e_t . Esta série é chamada de modelo de média móvel de ordem N , MA(N). As propriedades da série de saída x são claramente dependentes da série de entrada e . Para um sistema linear, a resposta do filtro é independente da entrada.

Num modelo **MA** têm-se três caracterizações equivalentes: no domínio do tempo (resposta de impulso de um filtro), no domínio da frequência (espectro), e em termos dos coeficientes das autocorrelações [48]. No primeiro caso, assume-se que a entrada é não zero somente num passo t_0 e é válido para todos os outros tempos. A resposta no domínio do tempo para este impulso é dado pela equação (1). A cada intervalo de tempo o impulso move acima para o próximo coeficiente até os N passos. As séries b_N, b_{N-1}, b_0 é o impulso de resposta do sistema. A resposta para uma entrada arbitrária pode ser calculada pela superposição das respostas sobre apropriados intervalos, pelos respectivos valores de entrada (convolução). A função de transferência assim completamente descreve um sistema linear onde a saída é determinada pela resposta de impulso e a entrada.

Se a entrada para o modelo **MA** é um impulso que possui um espectro de potência, a transformada discreta de fourier da saída é dada por:

$$\sum_{n=0}^N b_n \exp(-i2\pi n f)$$

O espectro de potência é dado pelo quadrado da magnitude da seguinte equação:

$$\left| 1 + b_1 e^{-i2\pi f} + b_2 e^{-i2\pi 2f} + \dots + b_N e^{-i2\pi Nf} \right|^2 \quad (2)$$

a terceira maneira de representar a mesma informação é em termos dos coeficientes da auto correlação, definidos em termos da média $\mu = \langle x_t \rangle$ e a variância $\sigma^2 = \langle (x_t - \mu)^2 \rangle$ dados pela seguinte expressão:

$$\rho_\tau \equiv \frac{1}{\sigma^2} \langle (x_t - \mu)(x_{t-\tau} - \mu) \rangle \quad (3)$$

O operador angular $\langle . \rangle$ denota valores esperados, na literatura estatística é indicado por $E\{. \}$.

Modelo auto regressivo (AR) O referido modelo opera em "open loop" com propagação para atrás; o modelo somente pode transformar uma entrada que é aplicada para tal. Se não se deseja manipular a série externamente, precisa-se prover alguma propagação para atrás em ordem para gerar dinâmica interna. A série representada pela equação:

$$x_t = \sum_{m=1}^M a_m x_{t-m} + e_t \quad (4)$$

é chamada de modelo auto regressivo de ordem M , (AR(M)) ou filtro de impulso de resposta infinito (IIR) por que a saída pode ser uma continuação da entrada [48]. Tal como o modelo anterior, se e_t é um ruído branco, as auto correlações da série de saída x pode ser expressas em termos dos coeficientes do modelo. Os coeficientes das autocorrelações de um modelo **AR** são encontrados resolvendo a equação linear, tradicionalmente chamado de equação de Yule-Walker.

$$\rho_\tau = \sum_{m=1}^M a_m \rho_{\tau-m} \quad \tau > 0 \quad (5)$$

Tomando a transformada de fourier de ambos os lados da equação (5) e rearranjando termos mostra que a saída é igual à entrada $(1 - \sum_{m=1}^M a_m \exp(-i2\pi mf))^{-1}$. O espectro de potência é dado pela seguinte expressão:

$$\frac{1}{\left| 1 - a_1 e^{-i2\pi f} - a_2 e^{-i2\pi 2f} - \dots - a_M e^{-2\pi Mf} \right|^2} \quad (6)$$

Para gerar uma realização específica das séries, precisa-se especificar condições iniciais, usualmente pelo primeiro valor \mathbf{M} da série \mathbf{x} .

O próximo passo em complexidade é permitir ambos modelos, ou seja, o modelo AR e o modelo MA num modelo só, este modelo é chamado de modelo ARMA(M,N) e escrito pela equação:

$$x_t = \sum_{m=1}^M a_m x_{t-m} + \sum_{n=0}^N b_n e_{t-n} \quad (7)$$

Esta saída é mais fácil de ser entendida em termos da transformada z , que generaliza a transformada discreta de fourier para o plano complexo:

$$X(z) \equiv \sum_{t=-\infty}^{\infty} x_t z^t$$

Dentro do círculo unitário, $z = \exp(-i2\pi f)$, a transformada z reduce-se para transformada discreta de fourier. Fora do círculo unitário, a transformada mede a taxa de convergência ou divergência da série. Desde que a convolução de dois séries no domínio do tempo correspondentes à multiplicação de suas transformadas z , a transformada z da saída de um modelo ARMA é:

$$X(z) = A(z)X(z) + B(z)E(z) = \frac{B(z)}{1 - A(z)} E(z) \quad (8)$$

A transformada $E(z)$ de entrada é multiplicada por uma função de transferência não relacionada com esta transformada; a função de transferência poderia convergir com zeros ($B(z)=0$) do termo MA e divergir com pólos ($A(z)=1$) do termo AR. Como $A(z)$ é um polinómio complexo de ordem M , e $B(z)$ de ordem N , poderia-se ter M pólos e N zeros. O número de pólos e zeros determina o número de graus de liberdade do sistema.

4.3.8 Ajuste do modelo linear para uma série temporal

A equação de *Yule-Walker* permite expressar os coeficientes de autocorrelação de uma série temporal em termos dos coeficientes AR gerados por este, assim como também permite estimar os coeficientes de um modelo AR(M). Um método alternativo é a estimação dos coeficientes através de regressão linear: expressando o próximo valor como uma função

dos valores prévios de M , exemplo: regressão linear de x_{t+1} dentro de $\{x_t, x_{t-1}, \dots, x_{t-(M-1)}\}$. Isto pode ser feito pela minimização dos erros quadrados: os parâmetros são determinados tal que a diferença entre os erros quadrados do modelo de saída e o valor observado seja minimizado. Esta simples expressão não é comparável conceitualmente para encontrar os coeficientes do modelo MA e ARMA dos dados observados. Para todos esses casos, existem técnicas standard, expressados como eficientes procedimentos recursivos.

4.3.9 Seleção do modelo

Uma questão de como estimar os coeficientes dos dados e a escolha do modelo ARMA de ordem (M, N) , ainda não tem sido tomado diretrizes [48]. Isto é, não há uma melhor escolha dos valores ou número de coeficientes para modelar uma série de dados. Várias heurísticas existem para encontrar a ordem correta do modelo, tais como, AIC (Akaike Information Criterion) e MDL (Minimum Description Length). Outras propostas de identificação tem sido apresentadas na literatura.

Akaike sugere escolher o modelo que minimiza o AIC (critério de informação de Akaike), que para modelos ARIMA (p, d, q) é dado por:

$$AIC(p, d, q) = N \log \sigma_a^2 + \frac{N}{N-d} 2(p + q + 1 + \sigma_{d0}) + N \log 2\pi + N$$

onde

$$\sigma_{d0} = \begin{cases} 1, & d=0 \\ 0, & d \neq 0 \end{cases}$$

A dificuldade da aplicação do método é que temos que ajustar um número grande de modelos. Por exemplo, se $0 \leq p \leq 2$, $0 \leq d \leq 2$, $0 \leq q \leq 2$, tem-se que estimar, ao todo, $3 \times 3 \times 3 = 27$ modelos e para cada um deles calcular o AIC. Um programa de computador desenvolvido por Akaike et al. chamado TIMSAC-78 contém programas especialmente desenvolvido para este fim.

4.3.10 Falha dos modelos lineares

Foi abordado que os coeficientes do modelo ARMA, o espectro de potência, e os coeficientes de auto correlação contêm a mesma informação de um sistema linear governado por ruído branco. Assim, se e somente se o espectro de potência descreve as características relevantes de uma série temporal, um modelo ARMA poderia ser uma boa escolha para descreve-la [48]. Esta característica pode falhar para sistemas não-lineares

quando estes geram espectros de potência complicados (ex: aproximadamente uniformemente distribuído).

Duas séries temporais podem ter espectro de potência similares porém podem ser geradas por sistemas com propriedades diferentes, tais como um sistema linear que é governado estocasticamente por um ruído externo, e um sistema linear determinístico com um pequeno número de graus de liberdade. Modelos não lineares são essenciais para a produção de interessantes comportamentos num sistema determinístico.

4.3.11 Modelo conexionista

Modelos conexionistas ou Redes Neurais não fazem uma descrição explícita de uma superfície a ser modelada, ou seja, elas prometem ser aplicáveis em situações onde métodos simples, ou mais explícitos falham. Muitas dessas promessas começam com a esperança que elas podem aprender a emular regularidades imprevistas num sinal complexo, ou seja, os modelos conexionistas são mais flexíveis que os modelos lineares.

A rede neural prevê dados futuros através de dados passados conhecidos. As arquiteturas são todas exemplos de redes do tipo feed-forward (propagação para frente). Para previsões um (ou poucos) passo(s) à frente elas são treinadas em "malha aberta", ou seja, utilizando dados medidos. Para previsões mais a longo prazo, a rede é treinada em "malha fechada". Neste caso são utilizados os valores previstos como entrada da rede durante o treinamento.

CAPÍTULO 5

Testes e Resultados

5.1 Introdução

Neste capítulo serão apresentados os resultados dos testes realizados para vários modelos de previsão, tanto de redes neurais como de métodos clássicos estatísticos. Dados de carga coletados pela Hydro-Quebec Power System do Canadá foram utilizados durante estes testes.

5.2 Teste 1: Comparação entre o algoritmo de retropropagação do erro e GMDH sem agrupamento de dados

Nesta comparação foi utilizada para treinamento uma base de dados de dois anos (Junho/Julho-1985 e 1986) e para teste os meses de Junho/Julho de 1987. Utilizando-se o algoritmo de retro propagação do erro foram testados diversas arquiteturas, com 6, 8, 10, 12 e 24 neurônios de entrada, encontrando-se como melhor arquitetura uma rede com doze (12) entradas, uma camada escondida com dez (10) neurônios e um (1) neurônio na camada de saída, como pode ser visto na figura 5.1.

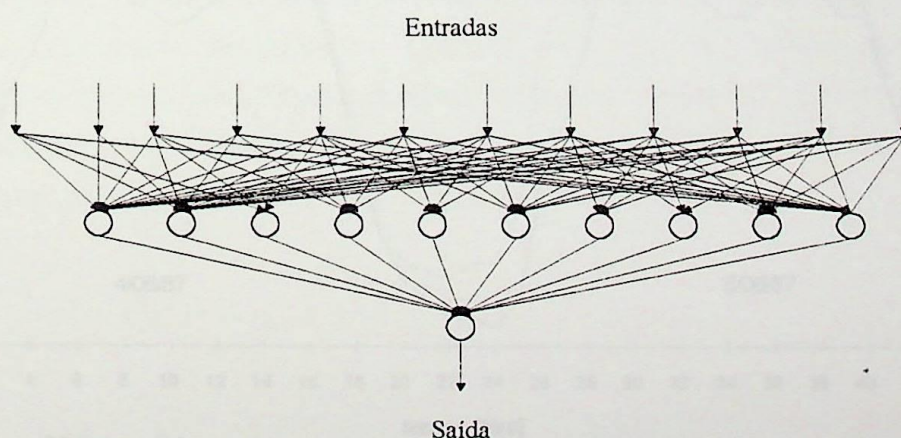


Figura 5.1 - Estrutura da rede neural (feed-forward) utilizada

Os erros de previsão um passo a frente (uma hora) para um período de 48 horas foram de:

Erro médio absoluto = 8,240 %

Erro máximo absoluto = 27,788 %

Erro mínimo absoluto = 0,013 %

Como pode ser observado na figura 5.2, os erros de previsão um passo a frente obtidos para essas horas foram muito elevados nos picos máximos e mínimos, acima dos 5% permitidos na prática. Esses erros são causados pelo uso de dados globais, isto é, dias da semana e fim de semana. Além disso, foram encontradas algumas dificuldades para se estabelecer a melhor configuração da rede para o processo de treinamento utilizado. A perfeita escolha do número de neurônios de cada camada é fundamental para uma boa solução do problema.

Para o caso do GMDH foram utilizados os mesmos dados acima para treinar e testar a rede. Os resultados obtidos na previsão foram de:

Erro médio absoluto = 3,564 %

Erro máximo absoluto = 6,106 %

Erro mínimo absoluto = 2,248 %

Como pode ser visto através dos resultados da figura 5.2, o GMDH que é uma rede neural não paramétrica onde a aproximação da curva é feita por polinômios, teve erros de previsão menores do que a rede feed-forward treinada via retropropagação do erro. Além disso, a curva de previsão encontra-se acima da curva medida, caracterizando previsões tendenciosas. No entanto a curva de previsão acompanha melhor os picos máximos e mínimos da curva medida.

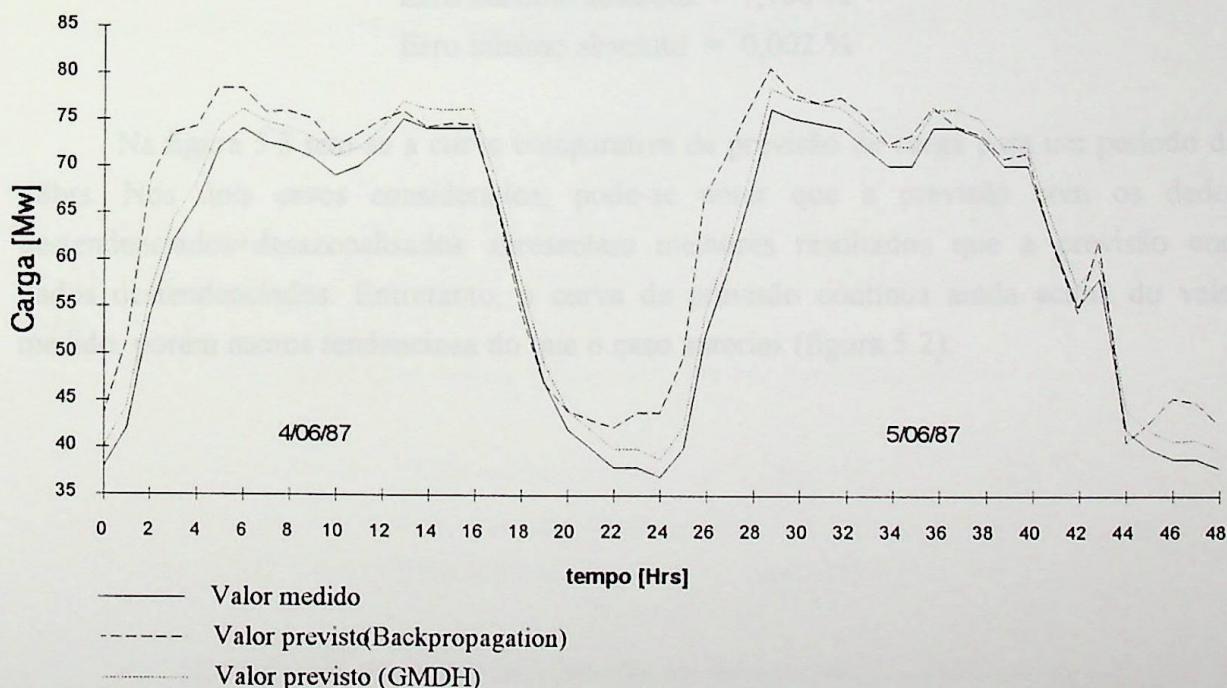


Figura 5.2 - Curva de previsão de carga comparativa para os dias 4/06/87 e 5/06/87 correspondente a 48 hrs, usando o algoritmo de retro propagação do erro e o GMDH com dados globais.



5.3 Teste 2: Comparação da previsão com dados globais destendenciados e destendenciados+desazonalisados, utilizando o GMDH

Para este caso foram utilizados os mesmos históricos de dados para o treinamento e teste do caso anterior. Primeiramente a série de dados foi destendenciada aplicando a primeira diferença, $Z_a = y(k) - y(k-1) = (1-B)y(k)$. Em seguida, os mesmos dados foram desazonalisados aplicando diferenciações sucessivas com atrasos iguais a 24 e 168, $Z_b = y(k) - y(k-24) = (1-B^{24})y(k)$ e $Z_c = y(k) - y(k-168) = (1-B^{168})y(k)$. Uma vez destendenciados e destendenciados+desazonalisados, os dados foram utilizados como entrada da rede. Obtendo-se os seguintes resultados:

Previsão com dados destendenciados

Erro médio absoluto = 2,725 %

Erro máximo absoluto = 4,327 %

Erro mínimo absoluto = 1,860 %

Previsão com dados destendenciados+desazonalisados

Erro médio absoluto = 0,544 %

Erro máximo absoluto = 7,780 %

Erro mínimo absoluto = 0,002 %

Na figura 5.3 tem-se a curva comparativa de previsão de carga para um período de 48hrs. Nos dois casos considerados, pode-se notar que a previsão com os dados destendenciados+desazonalisados apresentam melhores resultados que a previsão com dados destendenciados. Entretanto, a curva de previsão continua ainda acima do valor medido, porém menos tendenciosa do que o caso anterior (figura 5.2).

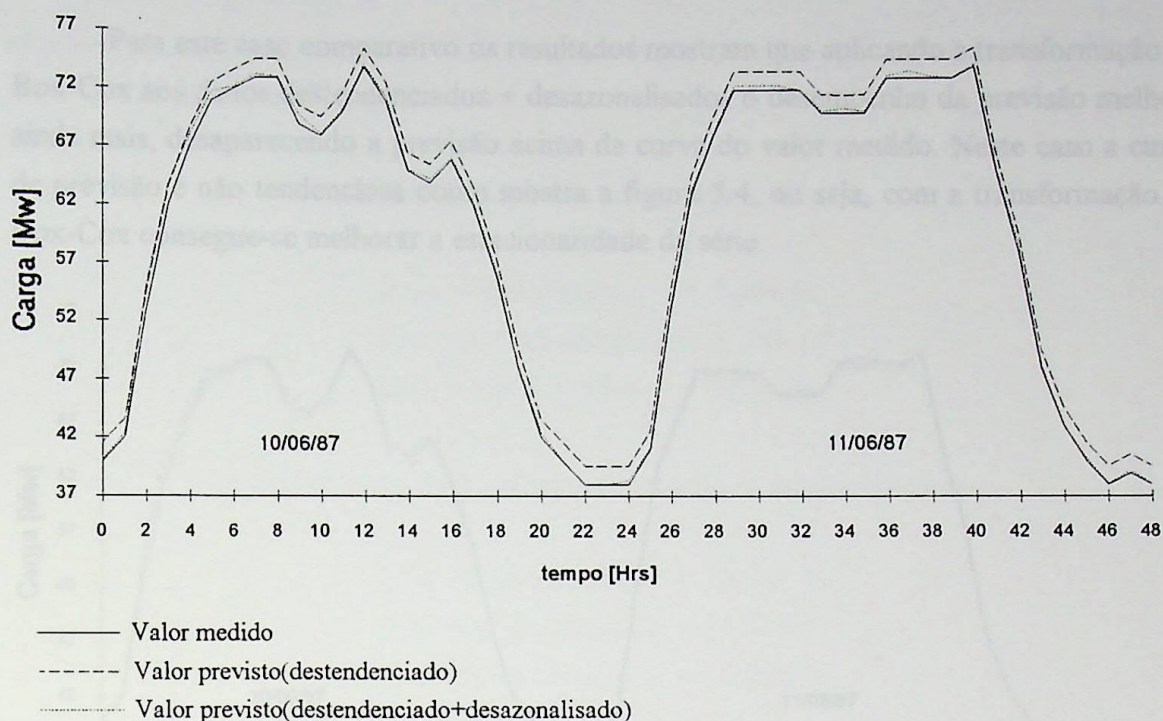


Figura 5.3 - Curva de previsão de carga comparativa para os dias 10/06/87 e 11/06/87 correspondentes a 48 hrs, utilizando dados globais destendenciados e destendenciados+desazonalizados

5.4 Teste 3: Comparação da previsão com dados globais destendenciados+desazonalizados e aplicando a transformação Box-Cox, usando o GMDH

Para este caso foram utilizados os mesmos dados de treinamento e teste dos casos anteriores, uma vez que os dados foram destendenciados e seguidamente desazonalizados foram utilizados como entrada na rede neural. Vale enfatizar que aplicando um atraso =168 hrs perde-se os dados de carga da primeira semana. Os resultados dos erros previstos foram:

GMDH com dados destendenciados+desazonalizados

Erro médio absoluto = 0,544 %

Erro máximo absoluto = 7,780 %

Erro mínimo absoluto = 0,002 %

GMDH com dados destendenciados+desazonalizados+Box-Cox

Erro médio absoluto = 0,946 %

Erro máximo absoluto = 2,289 %

Erro mínimo absoluto = 0,037 %

Para este caso comparativo os resultados mostram que aplicando a transformação de Box-Cox aos dados destendenciados + desazonalisados o desempenho da previsão melhora ainda mais, desaparecendo a previsão acima da curva do valor medido. Neste caso a curva de previsão é não tendenciosa como mostra a figura 5.4, ou seja, com a transformação de Box-Cox consegue-se melhorar a estacionaridade da série.

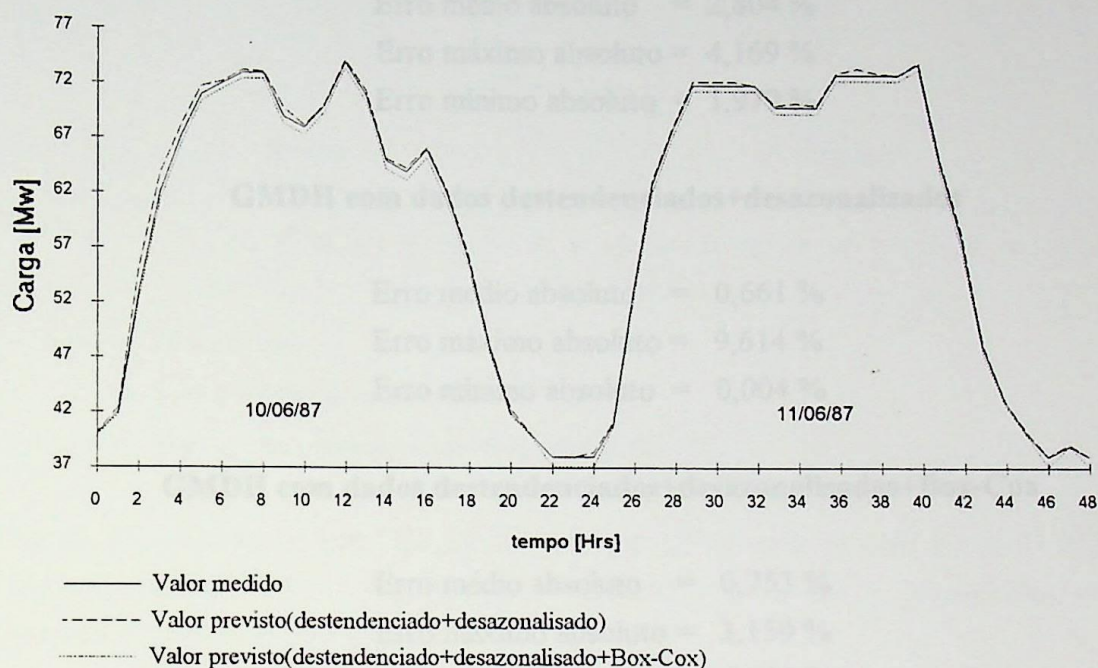


Figura 5.4 - Curva de previsão de carga comparativa para os dias 10/06/87 e 11/06/87 correspondentes a 48 hrs, utilizando dados destendenciados+desazonalisados e aplicando a transformação Box-Cox

5.5 Teste 4: Comparação da previsão com dados destendenciados, destendenciados+desazonalisados e aplicando a transformação Box-Cox, utilizando agrupamento de modelos de carga similares com o GMDH

Neste teste serem considerados três tipos de agrupamentos:

- Agrupamento dias sábados
- Agrupamento dias domingos
- Agrupamento dias da semana

5.5.1 Caso 1: Agrupamento para dias sábados

Nos dados agrupados para os dias sábados foram destendenciados e aplicados como entrada na rede neural (treinamento), seguidamente os mesmos dados agrupados foram desazonalisados e aplicado a transformação Box-Cox. Foram utilizados como treinamento

os meses de Junho e Julho de 1985 e 1986, e como teste os meses de Junho e Julho de 1987. Tendo como resultado o seguinte:

GMDH com dados destendenciados

Erro médio absoluto = 2,804 %

Erro máximo absoluto = 4,169 %

Erro mínimo absoluto = 1,979 %

GMDH com dados destendenciados+desazonalisados

Erro médio absoluto = 0,661 %

Erro máximo absoluto = 9,614 %

Erro mínimo absoluto = 0,004 %

GMDH com dados destendenciados+desazonalisados+Box-Cox

Erro médio absoluto = 0,753 %

Erro máximo absoluto = 3,159 %

Erro mínimo absoluto = 0,029 %

Pode ser notado através desses resultados que o agrupamento melhora o desempenho da previsão. Com dados destendenciados novamente pode ser visto através da figura 5.5 que a curva de previsão encontra-se acima da curva do valor medido, mais com dados destendenciados+desazonalisados diminui. Agora aplicando a transformação de Box-Cox os resultados melhoram em relação aos dados destendenciados+desazonalisados, não havendo aparecimento da previsão acima do valor medido. A figura 5.5 ilustra as curvas de previsão para 48 hrs com dados destendenciados e destendenciados+desazonalisados e a figura 5.6 ilustra a previsão também para 48 hrs com dados destendenciados+desazonalisados e aplicando a transformação de Box-Cox.

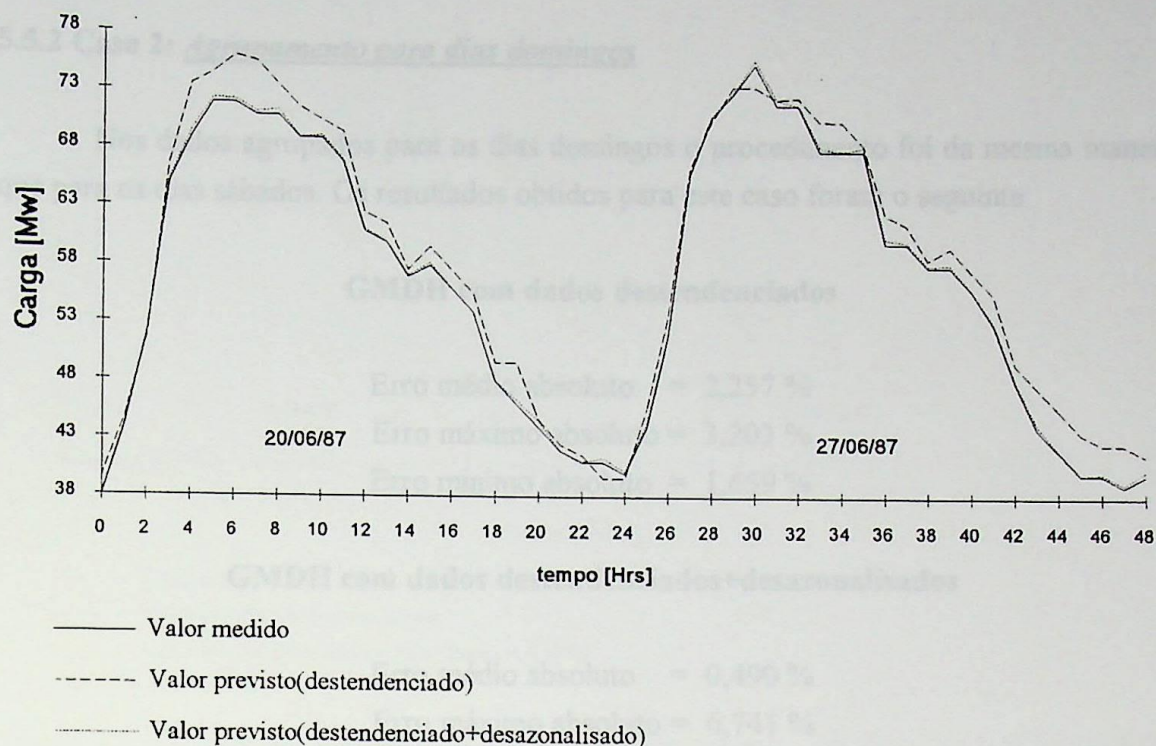


Figura 5.5 - Curva de previsão de carga comparativa para os dias 20/06/87 e 27/06/87 correspondentes a 48 hrs, utilizando dados destendenciados e destendenciados+desazonalisados

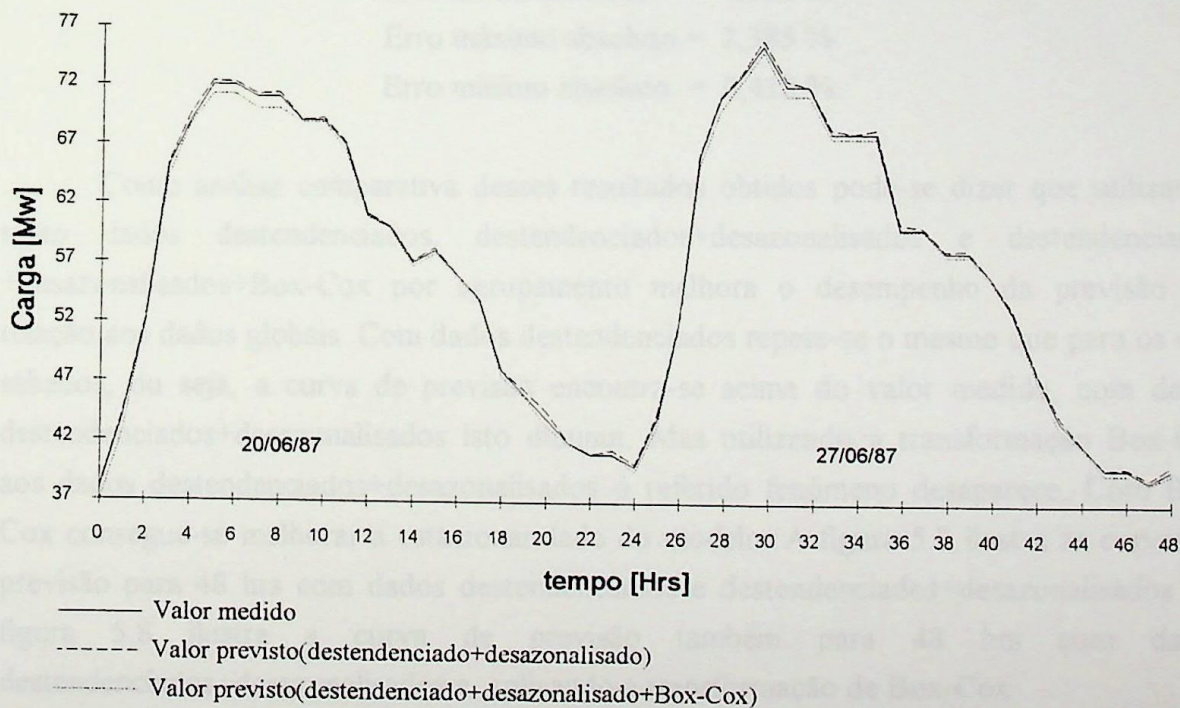


Figura 5.6 - Curva de previsão de carga comparativa para os dias 20/06/87 e 27/06/87 correspondentes a 48 hrs, utilizando dados destendenciados+desazonalisados e aplicando a transformação Box-Cox

5.5.2 Caso 2: Agrupamento para dias domingos

Nos dados agrupados para os dias domingos o procedimento foi da mesma maneira que para os dias sábados. Os resultados obtidos para este caso foram o seguinte:

GMDH com dados destendenciados

Erro médio absoluto = 2,257 %

Erro máximo absoluto = 3,203 %

Erro mínimo absoluto = 1,659 %

GMDH com dados destendenciados+desazonalisados

Erro médio absoluto = 0,490 %

Erro máximo absoluto = 6,741 %

Erro mínimo absoluto = 0,017 %

GMDH com dados destendenciados+desazonalisados+Box-Cox

Erro médio absoluto = 1,360 %

Erro máximo absoluto = 2,385 %

Erro mínimo absoluto = 0,412 %

Como análise comparativa desses resultados obtidos pode-se dizer que utilizando tanto dados destendenciados, destendenciados+desazonalisados e destendenciados+desazonalisados+Box-Cox por agrupamento melhora o desempenho da previsão em relação aos dados globais. Com dados destendenciados repete-se o mesmo que para os dias sábados, ou seja, a curva de previsão encontra-se acima do valor medido, com dados destendenciados+desazonalisados isto diminui. Mas utilizando a transformação Box-Cox aos dados destendenciados+desazonalisados o referido fenómeno desaparece. Com Box-Cox consegue-se melhorar a estacionaridade do modelo. A figura 5.7 ilustra as curvas de previsão para 48 hrs com dados destendenciados e destendenciados+desazonalisados e a figura 5.8 ilustra a curva de previsão também para 48 hrs com dados destendenciados+desazonalisados e aplicando a transformação de Box-Cox.

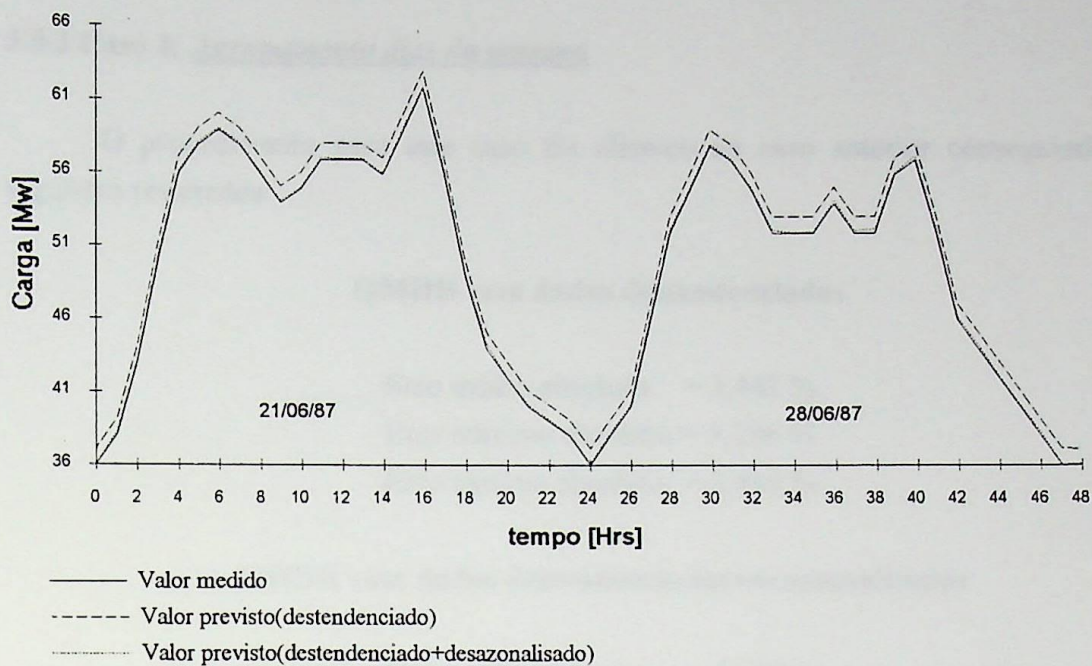


Figura 5.7 - Curva de previsão de carga comparativa para os dias 21/06/87 e 28/06/87 correspondentes a 48 hrs, utilizando dados destendenciados e destendenciados+desazonalisados

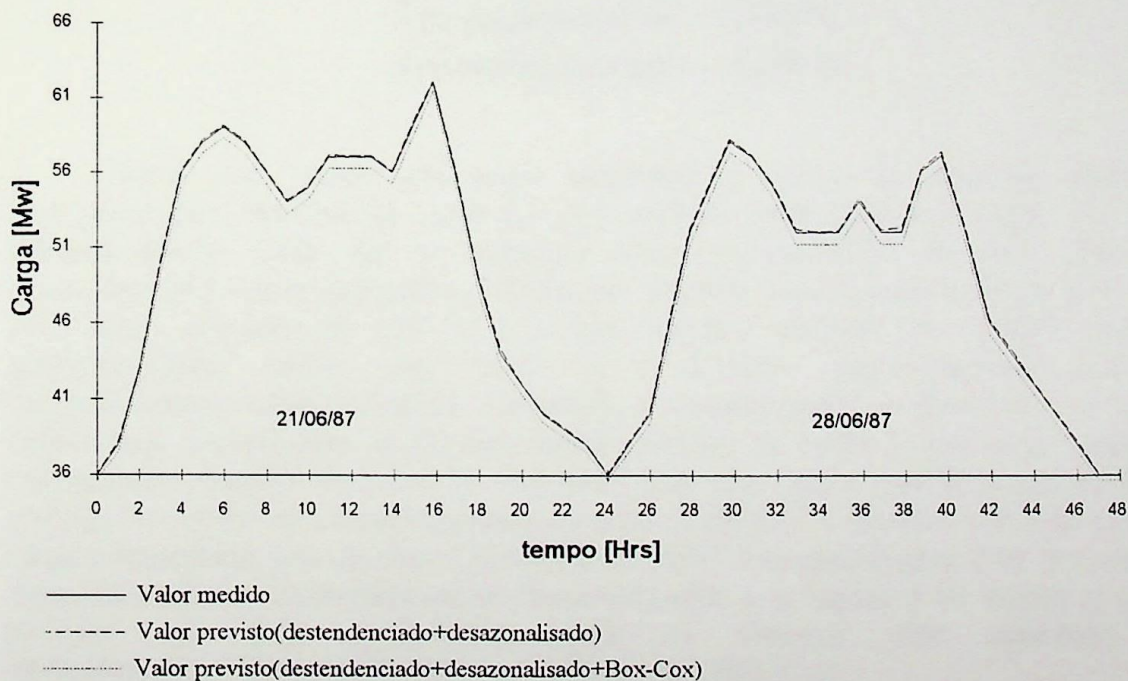


Figura 5.8 - Curva de previsão de carga comparativa para os dias 21/06/87 e 28/06/87 correspondentes a 48 hrs, utilizando dados destendenciados+desazonalisados e aplicando a transformação Box-Cox.

5.5.3 Caso 3: Agrupamento dias da semana

O procedimento para este caso foi idêntico ao caso anterior conseguindo-se os seguintes resultados:

GMDH com dados destendenciados

Erro médio absoluto = 3,442 %

Erro máximo absoluto = 5,296 %

Erro mínimo absoluto = 2,414 %

GMDH com dados destendenciados+desazonalisados

Erro médio absoluto = 0,637 %

Erro máximo absoluto = 4,116 %

Erro mínimo absoluto = 0,005 %

GMDH com dados destendenciados+desazonalisados+Box-Cox

Erro médio absoluto = 0,822 %

Erro máximo absoluto = 2,619 %

Erro mínimo absoluto = 0,004 %

Neste caso como em casos anteriores a curva de previsão com dados destendenciados encontra-se acima do valor medido como mostra a figura 5.9, ou seja, destendenciando ainda não se consegue uma estacionaridade da série. Com dados destendenciados+desazonalisados o fenômeno diminui como mostra a figura 5.10, onde o erro médio absoluto de 3,44 % e o erro máximo absoluto de 5,296% com dados destendenciados caíram para 0,637 % e 4,116% respectivamente com dados destendenciados+desazonalisados. Utilizando a transformação de Box-Cox os resultados melhoraram, conseguindo-se um erro médio absoluto de 0,822 % um pouco maior e um erro máximo absoluto de 2,619 %. Isto quer dizer que com a transformação de Box-Cox consegue-se melhorar a estacionaridade da série. A figura 5.9 ilustra a curva de previsão de carga comparativa para os dias 22/06/87 e 23/06/87 correspondentes a 48 hrs com dados destendenciados e destendenciados+desazonalisados e a figura 5.10 ilustra a curva de previsão de carga comparativa para os mesmos dias aplicando dados destendenciados+desazonalisados e a transformação Box-Cox.

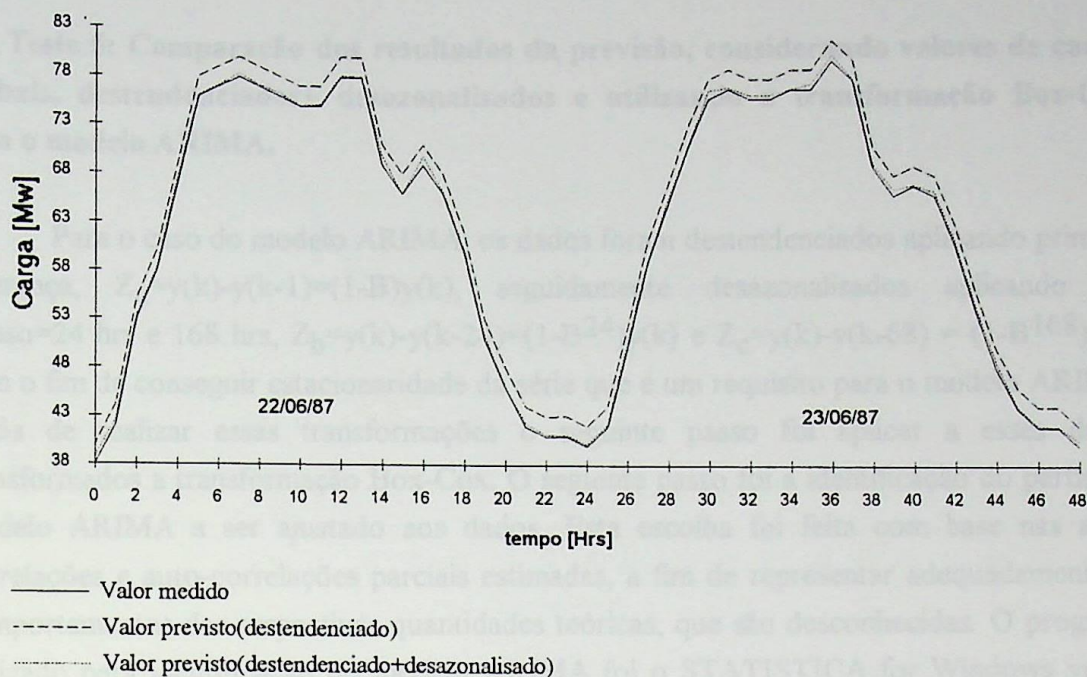


Figura 5.9 - Curva de previsão de carga comparativa para os dias 22/06/87 e 23/06/87 correspondentes a 48 hrs, utilizando dados destendenciados e destendenciados+desazonalisados

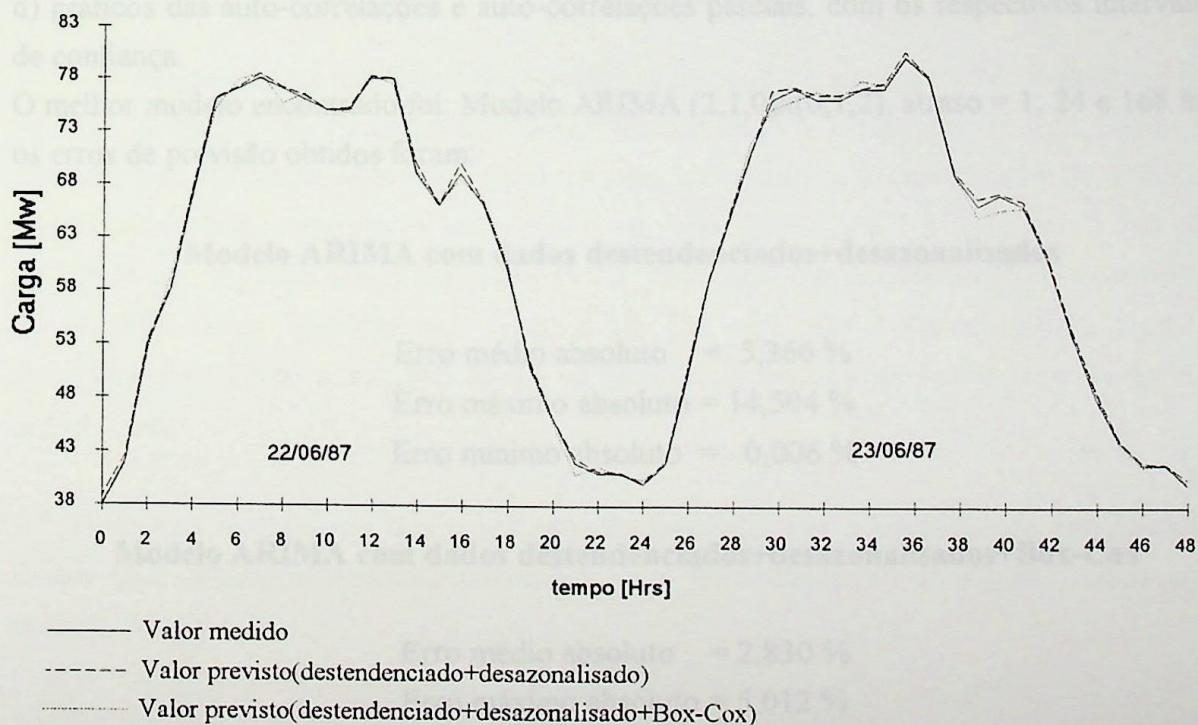


Figura 5.10 - Curva de previsão de carga comparativa para os dias 22/06/87 e 23/06/87 correspondentes a 48 hrs, utilizando dados destendenciados+desazonalisados e aplicando a transformação Box-Cox.

5.6 Teste 5: Comparação dos resultados da previsão, considerando valores de cargas globais, destendenciados+ desazonalizados e utilizando a transformação Box-Cox, com o modelo ARIMA.

Para o caso do modelo ARIMA, os dados foram destendenciados aplicando primeira diferença, $Z_a = y(k) - y(k-1) = (1-B)y(k)$, seguidamente desazonalizados aplicando um atraso=24 hrs e 168 hrs, $Z_b = y(k) - y(k-24) = (1-B^{24})y(k)$ e $Z_c = y(k) - y(k-68) = (1-B^{168})y(k)$ com o fim de conseguir estacionaridade da série que é um requisito para o modelo ARIMA. Após de realizar essas transformações o seguinte passo foi aplicar a esses dados transformados a transformação Box-Cox. O seguinte passo foi a identificação do particular modelo ARIMA a ser ajustado aos dados. Esta escolha foi feita com base nas auto-correlações e auto-correlações parciais estimadas, a fim de representar adequadamente os comportamentos das respectivas quantidades teóricas, que são desconhecidas. O programa utilizado para identificação do modelo ARIMA foi o STATISTICA for Windows versão 4.0. A saída de tal programa consta de:

- a) gráficos da série e das diferenças solicitadas;
- b) média e variância da série e de suas diferenças;
- c) auto-correlações e auto-correlações parciais estimadas;
- d) gráficos das auto-correlações e auto-correlações parciais, com os respectivos intervalos de confiança.

O melhor modelo encontrado foi: Modelo ARIMA (2,1,0)x(0,1,2), atraso = 1, 24 e 168 hrs, os erros de previsão obtidos foram:

Modelo ARIMA com dados destendenciados+desazonalizados

Erro médio absoluto = 5,366 %

Erro máximo absoluto = 14,504 %

Erro mínimo absoluto = 0,006 %

Modelo ARIMA com dados destendenciados+desazonalizados+Box-Cox

Erro médio absoluto = 2,830 %

Erro máximo absoluto = 5,012 %

Erro mínimo absoluto = 0,003 %

A figura 5.11 ilustra a curva de previsão comparativa para o modelo ARIMA, para os dias 20/06/87 e 21/06/87 correspondentes a 48 hrs.

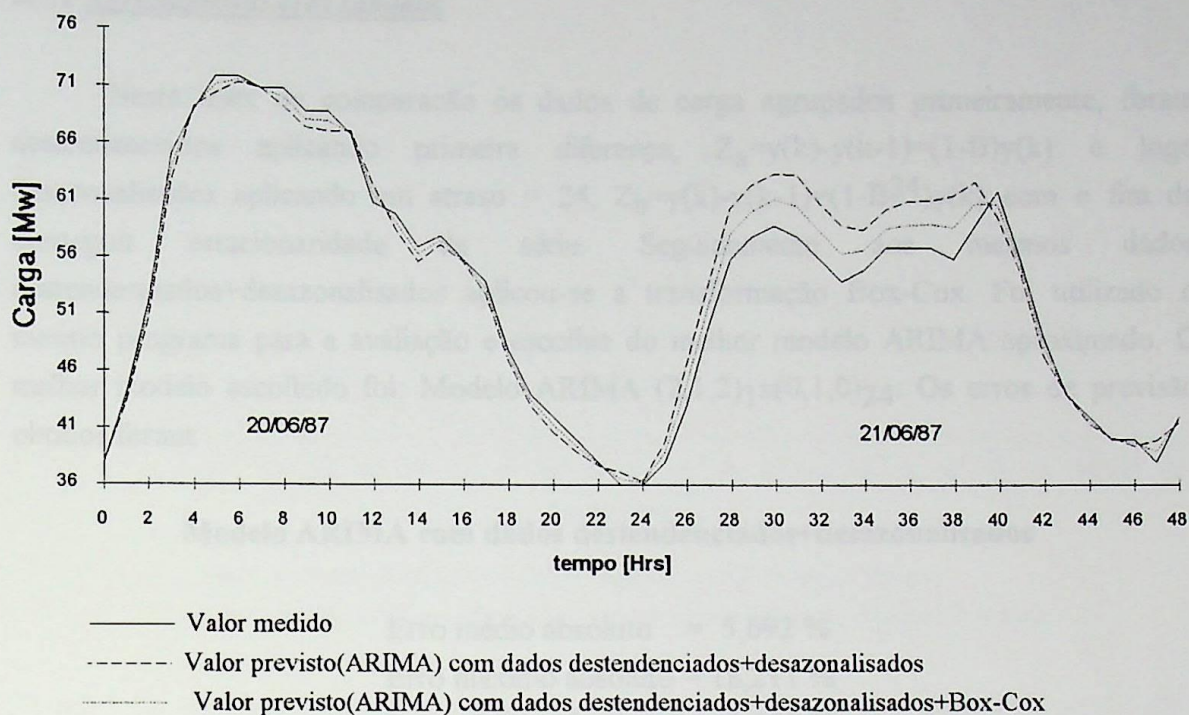


Figura 5.11- Curva de previsão de carga comparativa para os dias 20/06/87 e 21/06/87 correspondentes a 48 hrs com o modelo ARIMA, utilizando dados destendenciados+desazonalizados e a transformação de Box-Cox.

Comparando os resultados de previsão pode-se dizer que a previsão com o modelo ARIMA utilizando dados destendenciados+desazonalizados a margem de erro obtido foi elevado, caracterizando previsões tendenciosas. Utilizando a transformação de Box-Cox aos mesmos dados obteram-se erros muito menores 2,830 % e 5,017 % dos 5,366 % e 14,504 % obtidos com dados destendenciados+desazonalizados, melhorando o desempenho da previsão como ilustrado na figura 5.11. Ou seja com a transformação Box-Cox consegue-se melhorar a estacionaridade da série e conseguir uma melhor aproximação ao modelo real.

5.7 Teste 6: Comparação dos resultados da previsão por agrupamento considerando dados destendenciados+desazonalizados e a transformação Box-Cox, utilizando o modelo ARIMA

Nesta comparação serão considerados três tipos de agrupamentos.

- a) Agrupamento dias sábados
- b) Agrupamento dias domingos
- c) Agrupamento dias da semana

5.7.1 Agrupamento dias sábados

Neste teste de comparação os dados de carga agrupados primeiramente, foram destendenciados aplicando primeira diferença, $Z_a = y(k) - y(k-1) = (1-B)y(k)$ e logo desazonalisados aplicando um atraso = 24, $Z_b = y(k) - y(k-1) = (1-B^{24})y(k)$ com o fim de conseguir estacionaridade da série. Seguidamente aos mesmos dados destendenciados+desazonalisados aplicou-se a transformação Box-Cox. Foi utilizado o mesmo programa para a avaliação e escolha do melhor modelo ARIMA aproximado. O melhor modelo escolhido foi: Modelo ARIMA (2,1,2)₁x(0,1,0)₂₄. Os erros de previsão obtidos foram:

Modelo ARIMA com dados destendenciados+desazonalisados

Erro médio absoluto = 5,692 %

Erro máximo absoluto = 18,211 %

Erro mínimo absoluto = 0,152 %

Modelo ARIMA com dados destendenciados+desazonalisados+Box-Cox

Erro médio absoluto = 2,415 %

Erro máximo absoluto = 5,282 %

Erro mínimo absoluto = 0,127 %

A figura 5.12 ilustra a curva de previsão de carga comparativa para os dias 20/06/87, e 27/06/87 correspondentes a 48 hrs obtida com o modelo ARIMA. Através dos resultados dos erros obtidos pode ser visto que a previsão por agrupamento melhora em relação à previsão considerando cargas globais. Com dados destendenciados+desazonalisados os erros são muitos elevados. Agora utilizando a transformação Box-Cox aos mesmos dados destendenciados+desazonalisados o desempenho da previsão é melhorado, o erro médio absoluto de 5,692 % e o erro máximo absoluto de 18,211 % caíram para 2,415 % e 5,282 % respectivamente, ou seja, conseguiu-se melhorar o modelo.

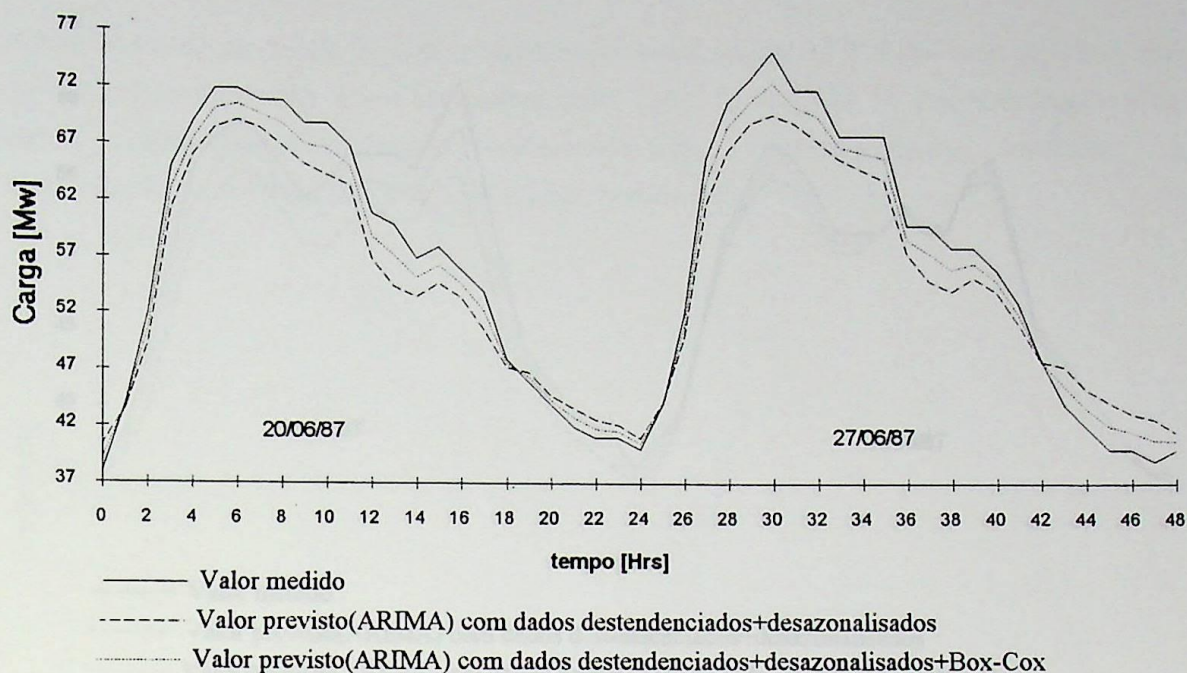


Figura 5.12 - Curva de previsão de carga comparativa para os dias 20/06/87, e 27/06/87 correspondente a 48 hrs obtida através do modelo ARIMA com dados destendenciados+desazonalisados e aplicando a transformação Box-Cox

5.7.2 Agrupamento dias domingos

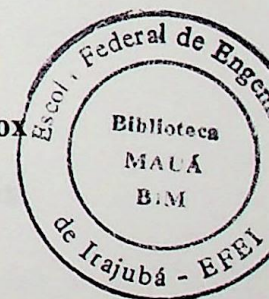
Para esta comparação foi utilizado os mesmos procedimentos do caso anterior. O modelo ARIMA escolhido para este caso foi: $(2,1,2)_1 \times (0,1,0)_{24}$ conseguindo-se os seguintes resultados:

Modelo ARIMA com dados destendenciados+desazonalisados

Erro médio absoluto = 5,101 %
 Erro máximo absoluto = 11,881 %
 Erro mínimo absoluto = 0,206 %

Modelo ARIMA com dados destendenciados+desazonalisados+Box-Cox

Erro médio absoluto = 2,501 %
 Erro máximo absoluto = 9,577 %
 Erro mínimo absoluto = 0,002 %



A figura 5.13 ilustra a curva de previsão de carga comparativa para os dias 21/06/87 e 28/06/87 correspondentes a 48hrs. Através da figura e os erros obtidos pode ser visto novamente a melhoria da previsão utilizando a transformação de Box-Cox aos dados destendenciados+desazonalisados. O modelo ARIMA sem a transformação Box-Cox apresenta previsões muito abaixo da curva de valor medido, agora com Box-Cox a previsão aproxima melhor a curva de valor medido.

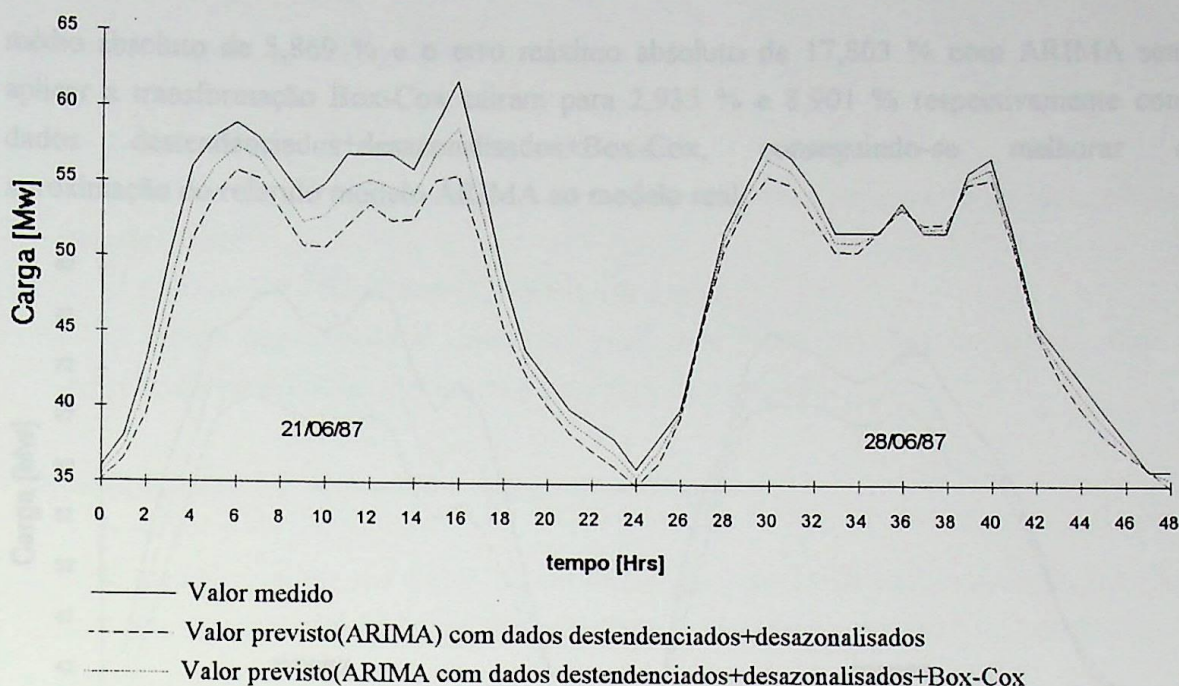


Figura 5.13 - Curva de previsão de carga comparativa para os dias 21/06/87 e 28/06/87 correspondentes a 48 Hrs obtida através do modelo ARIMA com dados destendenciados+desazonalisados e aplicando a transformação Box-Cox.

5.7.3 Agrupamento dias da Semana

Os mesmos procedimentos dos casos anteriores foram utilizados para este teste de comparação, o modelo ARIMA escolhido que melhor representa o modelo real foi: $ARIMA(2,1,2)_1 \times (0,1,0)_{24} \times (0,1,0)_{168}$ obtendo-se os seguintes erros:

Modelo ARIMA com dados destendenciados+desazonalisados

Erro médio absoluto = 5,869 %

Erro máximo absoluto = 17,803 %

Erro mínimo absoluto = 0,011 %

Modelo ARIMA com dados destendenciados+desazonalisados+Box-Cox

Erro médio absoluto = 2,935 %

Erro máximo absoluto = 8,901 %

Erro mínimo absoluto = 0,005 %

A figura 5.14 ilustra a curva de previsão de carga comparativa para os dias 22/06/87 e 23/06/87 correspondentes a 48 hrs. Através da figura e os resultados dos erros obtidos pode ser visto que os erros de previsão com o modelo ARIMA com dados destendenciados+desazonalisados são muito elevados, e ficam muito por abaixo da curva de valor medido. Com a transformação de Box-Cox obteve-se melhoria na previsão. O erro

médio absoluto de 5,869 % e o erro máximo absoluto de 17,803 % com ARIMA sem aplicar a transformação Box-Cox caíram para 2,935 % e 8,901 % respectivamente com dados destendenciados+desazonalisados+Box-Cox, conseguindo-se melhorar a aproximação do referido modelo ARIMA ao modelo real.

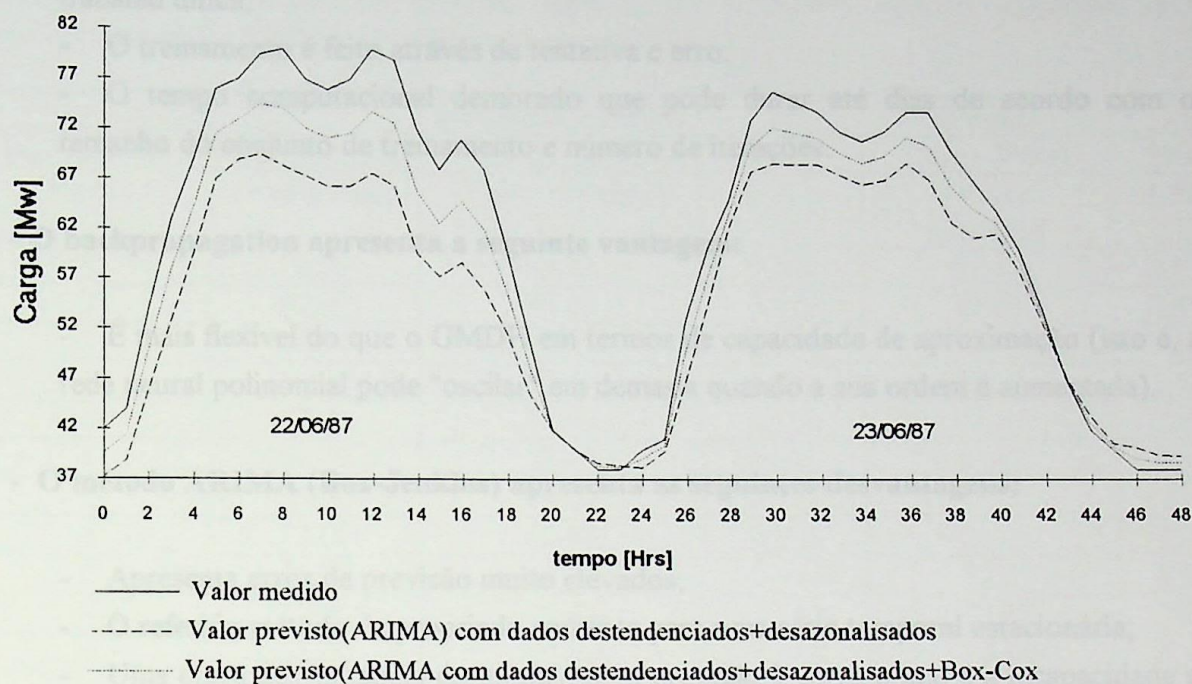


Figura 5.14 - Curva de previsão de carga comparativa para os dias 22/06/87 e 23/06/87 correspondentes a 48 hrs com o modelo ARIMA com dados destendenciados+desazonalisados e aplicando a transformação Box-Cox.

5.8 Conclusão dos testes realizados

Através da comparação dos testes realizados pode-se chegar às seguintes conclusões:

- O GMDH apresenta as seguintes vantagens:

- Apresenta menores erros de previsão em relação aos do backpropagation e do modelo ARIMA;
- Os erros com dados destendenciados e desazonalisados diminuem;
- Não precisa de parâmetros de ajuste do treinamento, e ainda cria a própria arquitetura;
- As variáveis não significativas são eliminadas pela rede reduzindo dessa maneira o maior número de variáveis e o tempo computacional;
- Os dados introduzidos na rede não precisam ser normalizados;

- As desvantagens do backpropagation são as seguintes:

- Apresenta erros de previsão muito elevados;
- Como se trata de uma rede paramétrica, escolher uma arquitetura apropriada é um trabalho difícil;
- O treinamento é feito através de tentativa e erro;
- O tempo computacional demorado que pode durar até dias de acordo com o tamanho do conjunto de treinamento e número de iterações.

- O backpropagation apresenta a seguinte vantagem:

- É mais flexível do que o GMDH em termos de capacidade de aproximação (isto é, a rede neural polinomial pode "oscilar" em demasia quando a sua ordem é aumentada).

- O método ARIMA (Box-Jenkins) apresenta as seguintes desvantagens:

- Apresenta erros de previsão muito elevados;
- O referido método é apropriado somente para uma série temporal estacionária;
- Uma outra desvantagem do modelo para previsão à curto prazo é a incapacidade de representar com precisão a relação não linear entre a carga e a temperatura;
- As previsões com o modelo ARIMA são essencialmente extrapolações dos históricos de carga e têm problemas quando ocorre mudanças imprevistas na temperatura;
- Outra dificuldade é a inclusão de variáveis independentes tais como: temperatura, velocidade do vento etc;
- Uma outra dificuldade na construção do modelo ARIMA, está na identificação, ou seja, em determinar os valores de p, d, q do modelo ARIMA (p, d, q) e o número de diferenças d necessárias para que o processo se torne estacionário.
- Finalmente, aplicando-se a transformação de Box-Cox aos mesmos dados destendenciados e desazonalizados melhora-se o desempenho da previsão tanto para redes neurais como para o modelo ARIMA.

Tabela 1

Resultados comparativos de previsão para o dia 01/06/87, com o algoritmo do Backpropagation e GMDH, considerando carga global

	destend.+desazon.			GMDH	
Dia 01/06/87	Carga [Mw]	Carga [Mw]	Erro	Carga [Mw]	Erro
Horas	[Valor real]	[Valor previsto]	[%]	[Valor previsto]	[%]
1	79,50	80,12	0,780	81,48	2,432
2	80,00	88,02	0,025	82,03	2,485
3	77,00	77,75	0,974	79,08	2,639
4	76,00	76,70	0,921	78,05	2,638
5	77,00	77,43	0,558	78,99	2,521
6	80,00	78,47	1,912	81,92	2,353
7	79,00	76,02	3,772	81,00	2,474
8	72,00	69,98	2,806	74,04	2,767
9	68,00	67,71	0,426	70,07	2,954
10	71,00	71,81	1,141	72,89	2,597
11	68,00	67,06	1,382	70,06	2,947
12	60,00	56,84	5,267	62,08	3,358
13	51,00	49,97	2,020	53,09	3,939
14	46,00	48,32	5,043	46,06	4,314
15	44,00	47,24	7,364	43,99	4,356
16	42,00	45,24	7,714	41,97	4,476
17	42,00	45,97	9,452	41,91	4,293
18	40,00	44,59	11,475	40,05	4,839
19	44,00	53,03	20,523	44,34	5,099
20	56,00	69,61	24,304	56,34	4,046
21	64,00	74,56	16,500	64,32	3,565
22	71,00	77,86	9,662	68,21	3,139
23	77,00	80,45	4,481	74,23	2,629
24	77,00	79,17	2,818	76,08	2,661
Erro médio absoluto			8,240 %		3,564 %
Erro máximo absoluto			27,788 %		6,106 %
Erro mínimo absoluto			0,013 %		2,248 %

Tabela 2

Resultados comparativos de previsão para o dia 10/06/87 com dados destendenciados, destendenciados+desazonalizados e destendenciados+desazonalizados+Box-Cox com dados globais, utilizando o GDMH.

Dia 01/06/87	Destendenciado		Destenden. + desazon.		Desten.+desazon.+Box-Cox		
	Carga [Mw] [Valor real]	Carga [Mw] [Valor previsto]	Erro [%]	Carga [Mw] [Valor previsto]	Erro [%]	Carga [Mw] [Valor previsto]	Erro [%]
1	40,0	41,577	3,793	40,110	0,274	39,89	0,275
2	42,0	43,586	3,639	42,410	0,967	41,75	0,595
3	53,0	54,605	2,939	55,180	3,951	52,28	1,358
4	62,0	63,611	2,533	63,790	2,806	61,22	1,258
5	67,0	68,593	2,322	68,000	1,471	66,31	1,030
6	71,0	72,589	2,189	71,780	1,087	70,44	0,789
7	72,0	73,592	2,163	72,180	0,249	71,48	0,722
8	73,0	74,613	2,134	73,260	0,355	72,52	0,658
9	73,0	70,599	2,162	73,130	0,179	72,52	0,658
10	69,0	69,608	2,265	69,850	1,217	68,37	0,913
11	68,0	71,595	2,310	68,180	0,264	67,34	0,971
12	70,0	75,601	2,228	70,140	0,200	69,40	0,857
13	74,0	72,597	2,118	74,230	0,310	73,57	0,581
14	71,0	66,597	2,200	71,410	0,574	70,44	0,789
15	65,0	65,569	2,398	65,170	0,261	64,27	1,123
16	64,0	67,576	2,393	64,400	0,621	63,25	1,172
17	66,0	63,594	2,332	66,200	0,302	65,29	1,076
18	62,0	57,592	2,507	62,190	0,306	61,22	1,258
19	56,0	49,605	2,764	56,400	0,709	55,23	1,375
20	48,0	43,594	3,236	48,011	0,023	47,44	1,167
21	42,0	41,603	3,656	42,410	0,967	41,75	0,595
22	40,0	39,602	3,876	40,002	0,005	39,89	0,275
23	38,0	39,597	4,045	38,011	0,029	37,13	2,289
24	38,0	39,583	4,033	38,002	0,005	37,13	2,289
Erro médio absoluto			2,725 %	0,544 %			0,946 %
Erro máximo absoluto			4,327 %	7,780 %			2,289 %
Erro mínimo absoluto			1,860 %	0,002 %			0,037 %

Tabela 3

Resultados comparativos de previsão por agrupamento para o dia sábado 20/06/87 com dados destendenciados, destendenciados+desazonalizados e destendenciados+desazonalizados+Box-Cox, utilizando o GDMH.

Dia 01/06/87	Destendenciado		Destenden. + desazon.		Desten.+desazon.+Box-Cox		
	Carga [Mw] [Valor real]	Carga [Mw] [Valor previsto]	Erro [%]	Carga [Mw] [Valor previsto]	Erro [%]	Carga [Mw] [Valor previsto]	Erro [%]
1	74,0	75,61	2,136	74,1	0,095	73,8	0,338
2	72,0	73,61	2,194	72,2	0,306	71,2	1,056
3	70,0	71,67	2,341	70,4	0,586	68,8	1,686
4	69,0	70,59	2,254	69,4	0,609	68,8	0,261
5	68,0	69,57	2,264	68,1	0,206	67,6	0,529
6	61,0	62,61	2,576	61,6	0,984	60,9	0,148
7	60,0	61,67	2,710	60,3	0,550	59,8	0,250
8	56,0	57,61	2,800	56,4	0,768	55,8	0,411
9	58,0	59,63	2,740	58,1	0,241	57,8	0,379
10	56,0	57,60	2,793	56,2	0,393	55,8	0,411
11	54,0	55,60	2,893	54,2	0,333	53,8	0,315
12	48,0	49,59	3,210	48,4	0,833	47,5	1,062
13	48,0	49,60	3,241	48,8	1,729	47,5	1,062
14	43,0	44,64	3,682	43,8	1,860	42,6	0,907
15	41,0	42,56	3,679	41,2	0,537	41,1	0,195
16	40,0	41,60	3,846	40,1	0,275	39,6	0,975
17	38,0	39,62	4,099	38,3	0,763	37,5	1,368
18	38,0	39,61	4,069	38,0	0,132	36,8	3,158
19	44,0	45,62	3,568	44,2	0,455	42,6	3,159
20	52,0	53,66	3,110	52,2	0,288	51,9	0,115
21	65,0	66,57	2,372	65,3	0,492	64,2	1,246
22	69,0	70,65	2,347	69,6	0,855	68,8	0,261
23	72,0	73,62	2,208	72,4	0,528	71,2	1,056
24	72,0	73,62	2,210	72,3	0,375	71,2	1,056
Erro médio absoluto			2,804 %	0,061 %			0,753 %
Erro máximo absoluto			4,169 %	9,614 %			3,159 %
Erro mínimo absoluto			1,979 %	0,004 %			0,029 %

Tabela 4

Resultados comparativos de previsão por agrupamento para o dia domingo 21/06/87 com dados destendenciados, destendenciados+desazonalizados e destendenciados+desazonalizados+Box-Cox, utilizando o GDMH.

Dia 01/06/87 Horas	Destendenciado		Destenden. + desazon.		Desten.+desazon.+Box-Cox		
	Carga [Mw] [Valor real]	Carga [Mw] [Valor previsto]	Erro [%]	Carga [Mw] [Valor previsto]	Erro [%]	Carga [Mw] [Valor previsto]	Erro [%]
1	62,0	63,128	1,819	62,248	0,400	61,42	0,935
2	59,0	60,124	1,905	59,149	0,253	58,31	1,169
3	56,0	57,125	2,009	56,151	0,270	55,23	1,375
4	55,0	56,128	2,051	55,010	0,018	54,22	1,418
5	57,0	58,138	1,996	57,100	0,175	56,25	1,316
6	58,0	59,126	1,941	58,104	0,179	57,28	1,241
7	58,0	59,124	1,938	58,149	0,257	57,28	1,241
8	59,0	60,129	1,914	59,248	0,420	58,31	1,169
9	61,0	62,131	1,854	61,100	0,164	60,38	1,016
10	62,0	63,126	1,816	62,149	0,240	61,42	0,935
11	57,0	58,121	1,967	57,149	0,261	56,25	1,316
12	48,0	49,123	2,340	48,106	0,221	47,20	1,667
13	44,0	45,130	2,568	44,348	0,791	43,29	1,614
14	42,0	43,130	2,690	42,104	0,248	41,35	1,548
15	40,0	41,129	2,823	40,202	0,505	39,44	1,400
16	38,0	39,129	2,971	38,249	0,655	37,54	1,211
17	36,0	37,132	3,144	36,104	0,289	35,66	0,944
18	36,0	37,119	3,108	36,054	0,150	35,66	0,944
19	38,0	39,119	2,945	38,248	0,653	37,54	1,211
20	43,0	44,132	2,633	43,054	0,126	42,32	1,581
21	50,0	51,124	2,248	50,298	0,596	49,19	1,620
22	56,0	57,128	2,014	56,149	0,266	55,23	1,375
23	58,0	59,127	1,943	58,151	0,260	57,28	1,241
24	59,0	60,127	1,943	59,104	0,176	58,31	1,169
Erro médio absoluto			2,257 %	0,490 %			1,360 %
Erro máximo absoluto			3,203 %	6,741 %			2,385 %
Erro mínimo absoluto			1,659 %	0,017 %			0,412 %

Tabela 5

Resultados comparativos de previsão por agrupamento para o dia de semana 22/06/87 com dados destendenciados, destendenciados+desazonalizados e destendenciados+desazonalizados+Box-Cox, utilizando o GDMH.

Dia 01/06/87	Destendenciado		Destenden. + desazon.		Desten.+desazon.+Box-Cox		
	Carga [Mw] [Valor real]	Carga [Mw] [Valor previsto]	Erro [%]	Carga [Mw] [Valor previsto]	Erro [%]	Carga [Mw] [Valor previsto]	Erro [%]
1	38,0	40,077	5,183	38,770	2,026	37,830	0,447
2	42,0	44,069	4,695	42,550	1,310	41,900	0,238
3	53,0	55,073	3,764	53,350	0,660	52,400	1,132
4	58,0	60,084	3,468	58,270	0,466	59,300	2,241
5	67,0	69,081	3,012	67,730	1,090	67,440	0,657
6	76,0	78,076	2,659	76,110	0,145	75,150	1,118
7	77,0	79,080	2,630	77,160	0,208	77,810	1,052
8	78,0	80,069	2,584	78,420	0,538	78,420	0,538
9	77,0	79,063	2,609	77,100	0,130	75,620	1,792
10	76,0	78,087	2,673	76,360	0,474	76,490	0,645
11	75,0	77,065	2,680	75,030	0,040	74,930	0,093
12	75,0	77,070	2,686	75,190	0,253	74,770	0,307
13	78,0	80,071	2,594	78,220	0,282	78,300	0,385
14	78,0	80,074	2,590	78,040	0,051	77,410	0,756
15	69,0	71,076	2,921	69,940	1,362	69,230	0,333
16	66,0	68,078	3,052	66,041	0,062	66,290	0,439
17	69,0	71,076	2,921	69,950	1,377	68,400	0,870
18	66,0	68,080	3,055	66,270	0,409	66,350	0,530
19	60,0	62,075	3,343	60,260	0,433	60,760	1,267
20	51,0	53,078	3,915	51,430	0,843	50,780	0,431
21	46,0	48,069	4,304	46,021	0,046	45,720	0,630
22	42,0	44,083	4,725	42,370	0,881	40,900	2,619
23	41,0	43,070	4,806	41,200	0,488	41,620	1,512
24	41,0	43,090	4,850	41,080	0,195	40,890	0,268
Erro médio absoluto			3,442 %		0,637 %		0,822 %
Erro máximo absoluto			5,296 %		4,116 %		2,619 %
Erro mínimo absoluto			2,414 %		0,005 %		0,004 %

Tabela 6

Resultados comparativos de previsão para o dia 20/06/87 com dados globais, destendenciados+desazonalizados e destendenciados+desazonalizados+Box-Cox, utilizando o modelo ARIMA.

Dia 01/06/87	Carga [Mw]		Destenden. + desazon.		Desten.+desazon.+Box-Cox	
	[Valor real]	[Valor previsto]	[Valor previsto]	Erro [%]	[Valor previsto]	Erro [%]
Horas						
1	43,0	41,0	41,0	4,749	42,0	2,374
2	41,0	39,0	39,0	4,841	40,0	2,422
3	40,0	38,3	38,3	4,288	39,1	2,142
4	38,0	37,7	37,7	0,747	37,9	0,374
5	38,0	38,0	38,0	0,029	38,0	0,016
6	44,0	43,3	43,3	1,695	43,6	0,848
7	52,0	50,6	50,6	2,783	51,3	1,390
8	65,0	62,8	62,8	3,442	63,9	1,720
9	69,0	69,3	69,3	0,491	69,2	0,245
10	72,0	70,5	70,5	2,081	71,3	1,040
11	72,0	71,5	71,5	0,742	71,7	0,371
12	71,0	71,0	71,0	0,006	71,0	0,003
13	71,0	69,4	69,4	2,294	70,2	1,146
14	69,0	67,5	67,5	2,123	68,3	1,061
15	69,0	67,2	67,2	2,568	68,1	1,284
16	67,0	67,3	67,3	0,397	67,1	0,199
17	61,0	62,5	62,5	2,530	61,8	1,264
18	60,0	58,6	58,6	2,268	59,3	1,135
19	57,0	55,7	55,7	2,282	56,3	1,140
20	58,0	57,2	57,2	1,312	57,6	0,657
21	56,0	55,7	55,7	0,493	55,9	0,246
22	54,0	52,9	52,9	2,098	53,4	1,048
23	48,0	47,0	47,0	2,019	47,5	1,010
24	44,0	43,0	43,0	2,291	43,5	1,145
Erro médio absoluto				5,366 %		2,830 %
Erro máximo absoluto				14,504 %		5,012 %
Erro mínimo absoluto				0,006 %		0,003 %

Tabela 7

Resultados comparativos de previsão por agrupamento para o dia sábado 20/06/87 com dados destendenciados+desazonalizados e destendenciados+desazonalizados+Box-Cox, utilizando o modelo ARIMA.

Dia 01/06/87	Carga [Mw]		Destenden. + desazon.		Desten.+desazon.+Box-Cox	
	[Valor real]	[Valor previsto]	Carga [Mw]	Erro [%]	Carga [Mw]	Erro [%]
Horas			[Valor previsto]		[Valor previsto]	
1	48,0	46,2	46,2	3,688	47,1	1,844
2	43,0	44,1	44,1	2,670	43,6	1,335
3	41,0	43,0	43,0	4,963	42,0	2,483
4	40,0	42,0	42,0	4,983	41,0	2,490
5	38,0	41,6	41,6	9,395	39,8	4,697
6	38,0	40,4	40,4	6,250	39,2	3,126
7	44,0	43,5	43,5	1,089	43,8	0,545
8	52,0	49,3	49,3	5,100	50,7	2,550
9	65,0	61,3	61,3	5,654	63,2	2,828
10	69,0	66,1	66,1	4,238	67,5	2,119
11	72,0	68,5	68,5	4,812	70,3	2,406
12	72,0	69,2	69,2	3,822	70,6	1,911
13	71,0	68,5	68,5	3,458	69,8	1,730
14	71,0	66,9	66,9	5,730	69,0	2,865
15	69,0	65,4	65,4	5,259	67,2	2,629
16	69,0	64,4	64,4	6,700	66,7	3,351
17	67,0	63,5	63,5	5,248	65,2	2,624
18	61,0	56,9	56,9	6,733	58,9	3,366
19	60,0	54,5	54,5	9,190	57,2	4,595
20	57,0	53,6	53,6	5,996	55,3	2,998
21	58,0	54,8	54,8	5,598	56,4	2,800
22	56,0	53,5	53,5	4,455	54,8	2,227
23	54,0	50,6	50,6	6,285	52,3	3,143
24	48,0	47,3	47,3	1,454	47,7	0,727
Erro médio absoluto				5,692 %		2,415 %
Erro máximo absoluto				18,211 %		5,282 %
Erro mínimo absoluto				0,152 %		0,127 %

Tabela 8

Resultados comparativos de previsão por agrupamento para o dia domingo 21/06/87 com dados destendenciados+desazonalizados e destendenciados+desazonalizados+Box-Cox, utilizando o modelo ARIMA.

Dia 01/06/87		Destenden. + desazon.		Desten.+desazon.+Box-Cox	
Horas	Carga [Mw] [Valor real]	Carga [Mw] [Valor previsto]	Erro [%]	Carga [Mw] [Valor previsto]	Erro [%]
1	44,0	42,4	3,691	44,2	0,343
2	42,0	40,2	4,245	40,8	2,850
3	40,0	38,4	3,985	39,1	2,275
4	38,0	37,2	2,068	36,9	2,813
5	36,0	36,2	0,481	35,9	0,214
6	36,0	34,9	3,064	37,5	4,033
7	38,0	36,4	4,158	40,3	5,989
8	43,0	39,4	8,272	47,1	9,577
9	50,0	45,5	8,944	53,1	6,278
10	56,0	50,8	9,205	56,7	1,211
11	58,0	53,8	7,221	58,3	0,557
12	59,0	55,7	5,612	57,8	2,083
13	58,0	55,2	4,876	57,3	1,247
14	56,0	53,0	5,411	54,8	2,116
15	54,0	50,7	6,076	54,9	1,674
16	55,0	50,7	7,909	56,0	1,791
17	57,0	52,1	8,565	57,8	1,432
18	57,0	53,5	6,088	56,7	0,607
19	57,0	52,5	7,902	57,1	0,195
20	56,0	52,6	6,059	57,5	2,768
21	59,0	55,2	6,415	59,8	1,305
22	62,0	55,6	10,398	58,7	5,335
23	56,0	51,1	8,695	52,7	5,838
24	49,0	45,4	7,310	46,6	4,837
Erro médio absoluto			5,101 %	2,501 %	
Erro máximo absoluto			11,881 %	9,577 %	
Erro mínimo absoluto			0,206 %	0,002 %	

Tabela 9

Resultados comparativos de previsão por agrupamento para o dia da semana 22/06/87 com dados destendenciados+desazonalizados e destendenciados+desazonalizados+Box-Cox, utilizando o modelo ARIMA.

Dia 01/06/87	Carga [Mw]		Destenden. + desazon.		Desten.+desazon.+Box-Cox	
	[Valor real]	[Valor previsto]	Carga [Mw]	Erro [%]	Carga [Mw]	Erro [%]
Horas						
1	44,0	41,5	42,7	5,730	42,7	2,864
2	42,0	38,7	40,4	7,779	40,4	3,888
3	40,0	38,1	39,1	4,657	39,1	2,327
4	40,0	37,4	38,7	6,415	38,7	3,208
5	42,0	37,3	39,6	11,295	39,6	5,648
6	44,0	38,9	41,4	11,691	41,4	5,845
7	54,0	45,9	50,0	14,928	50,0	7,465
8	63,0	52,9	58,0	15,960	58,0	7,981
9	69,0	59,5	64,3	13,765	64,3	6,883
10	76,0	66,9	71,5	11,934	71,5	5,967
11	77,0	68,9	73,0	10,457	73,0	5,229
12	79,5	69,5	74,5	12,555	74,5	6,278
13	80,0	68,3	74,2	14,613	74,2	7,306
14	77,0	67,3	72,1	12,622	72,1	6,310
15	76,0	66,3	71,1	12,800	71,1	6,400
16	77,0	66,2	71,6	13,971	71,6	6,986
17	80,0	67,6	73,8	15,464	73,8	7,732
18	79,0	66,3	72,7	16,025	72,7	8,013
19	72,0	59,5	65,7	17,390	65,7	8,694
20	68,0	57,3	62,6	15,800	62,6	7,900
21	71,0	58,7	64,9	17,315	64,9	8,658
22	68,0	55,9	61,9	17,803	61,9	8,901
23	60,0	51,6	55,8	13,947	55,8	6,973
24	51,0	45,5	48,3	10,737	48,3	5,369
Erro médio absoluto				5,869 %		2,935 %
Erro máximo absoluto				17,803 %		8,901 %
Erro mínimo absoluto				0,011 %		0,005 %

CAPÍTULO 6

CONCLUSÕES

Neste trabalho foi apresentado uma análise comparativa de métodos de previsão de carga, utilizando Redes Neurais Artificiais e métodos estatísticos clássicos (Box & Jenkins).

Primeiramente foi desenvolvido um programa baseado em distâncias Euclidianas para a identificação de dias típicos, ou seja, agrupamentos de modelos de curvas de cargas com características similares.

Para conseguir agrupar os modelos de carga, os dados de carga tiveram que ser destendenciados. Foram obtidos três agrupamentos de carga: sábados, domingos e dias de semana.

Dos resultados obtidos através das simulações, pode-se chegar às seguintes conclusões:

1. Fazendo previsão com redes neurais considerando a carga não agrupada (dias da semana e fim de semana conjuntamente), a margem de erro obtida foi muito elevada. Isto é devido à não similaridade dos padrões de curva de carga.
2. Considerando agrupamento de dados de carga, os erros de previsão obtidos foram melhorados.
3. Destendenciando os dados de carga os erros previstos diminuíram, embora as previsões tenham sido tendenciosas.
4. Destendenciando e Desazonalizando os dados, as previsões foram melhoradas, embora permanecendo tendenciosas.
5. Aplicando a transformação de Box-Cox aos dados destendenciados e desazonalizados os erros de previsão são diminuídos ainda mais. Além disso, as previsões deixam de ser tendenciosas, não somente para o caso de redes neurais assim como para o modelo

ARIMA. Ou seja com a transformação de Box-Cox colabora efetivamente para "estacionar" a série de carga.

6. Na previsão com o agrupamento os erros previstos obtidos foram menores, pois os modelos de carga tornam-se mais simples.
7. Na previsão utilizando-se redes neurais, o GMDH teve melhor desempenho que o algoritmo de retro-propagação do erro (backpropagation).

Em geral podemos dizer que as redes neurais apresentaram melhor desempenho em relação aos métodos clássicos, ou seja elas tem a capacidade de modelar sistemas complexos e produzir excelentes resultados. Isto poderia certamente representar um novo horizonte na pesquisa, não somente na previsão de carga, mais sim em outras áreas de natureza similar. A previsão por agrupamento apresentou bons resultados, tanto para as redes neurais como para o modelo ARIMA.

Proposta de futuros trabalhos

Tendo considerado neste trabalho a carga como única variável, outras variáveis como temperatura e velocidade do vento poderiam ser introduzidas na rede neural e verificar se a introdução dessas variáveis afetariam na precisão da previsão, uma outra proposta seria a utilização do modelo da expansão polinomial e verificar se o referido modelo afetaria ou não no aprendizado da rede. Outros modelos e algoritmos poderiam ser testados experimentalmente.

[10] J. W. S. CHRISTIANINI - "Short-term load forecasting using General Exponential Smoothing", IEEE Trans Power App Syst., Vol. PAS-90, No. 2, pp. 900-911, Mar/Apr 1971.

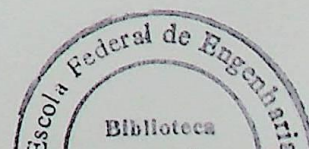
[11] L. BIANCHI, J. RODRIGUEZ - "Adaptive forecasting of hourly load based on load measurements and weather information", IEEE Trans Power App Syst., Vol. PAS-90, No. 4, pp. 1757-1767, July/Aug. 1971.

[12] D. NIEBUR, G. LAMBERT TORRES, A.P. ALVES DA SILVA, et al. - "Artificial Neural Networks for Power Systems", Eletric. No. 139, pp. 76-101, April 1995.

[13] G. LAMBERT TORRES & A.P. ALVES DA SILVA - "Aplicação de Redes Neurais Inteligentes em Engenharia de Potência", Controle e Automação Vol. 8, pp. 10-15, Oct./Nov. 1994.

REFERÊNCIAS

- [1] G.GROSS & F.D. GALIANA, - " Short term load forecasting", Proc. IEEE, Vol. 75, No. 12, pp. 1558-1573, 1987.
- [2] G. LAMBERT TORRES, L.E. BORGES DA SILVA, B. VALIQUETTE, H. GREISS & D. MUKHEDKAR, - "A fuzzy Knowledge-Based System for Bus Load Forecasting", Fuzzy Logic Technology and Applications, por Robert J. Marks II, IEEE Press, pp. 221-228, ISBN 0-7803-1383-6, 1994.
- [3] P.C. GUPTA, - "A stochastic approach to peak power demand forecasting in electric utility systems", IEEE. Trans. Power. App. Syst., Vol. PAS-90, Mar./Apr. 1971.
- [4] T.N. GOH, H.L. ONG & Y.O. LEE, - "A new approach to statistical forecasting of daily peak power demand", Elec. Power Syst. Res., Vol. 10, No. 2, pp. 145-148, Mar. 1986.
- [5] S. RAHMAN & R. BHATNAGAR, - "An expert system based algorithm for short term load forecasting", IEEE. Trans. Power Systems., Vol. 1, pp 392-398. 1988.
- [6] G. LAMBERT TORRES, B. VALIQUETTE & D. MUKHEDKAR, - "Bus Load Forecasting Using Fuzzy Techniques", CEA Engineering and Operating Transactions, Vol. 29, Part IV, 1990.
- [7] T.S. DILLON, K. MORSZTYN & K. PHUA, - "Short term Load Forecasting Using Adaptive Pattern Recognition and Self Organizing Techniques", "Proc. 5th PSCC, Cambridge, Vol. 1, paper 2.4/3, Sept. 1975.
- [8] A.S. DEHDASHTI, J.R. TUDOR, & M.C. Smith, - "Forecasting of hourly load by Pattern Recognition-A Deterministic Approach", IEEE. Trans. Power App. Syst., Vol. PAS-101, Vol. 9, pp. 3290-3294, Sept. 1982.
- ✓ [9] W.R. CHRISTIAANSE, - "Short-term load forecasting using General Exponential Smoothing", IEEE. Trans. Power App. Syst., Vol. PAS-90, No.2, pp. 900-911, Mar./Apr. 1971.
- [10] LIJESSEN, J. ROSING, - "Adaptive forecasting of hourly load based on load measurements and weather information", IEEE Trans. Power App. Syst., Vol. PAS-90, No. 4, pp. 1757-1767, July/Aug. 1971.
- [11] D.NIEBUR, G. LAMBERT TORRES, A.P. ALVES DA SILVA, et al. - "Artificial Neural Network for Power Systems", Electra, No. 159, pp. 76-101, April 1995.
- [12] G. LAMBERT TORRES & A.P. ALVES DA SILVA - "Aplicação de Sistemas Inteligentes em Engenharia de Potência", Controle e Automação, Vol. 5, pp. 144-152, Out./Nov. 1994.



- [13] I. MOGHARAM, S. RAHMAN, - "Analysis and Evaluation of Five Short-Term Load forecasting Techniques", IEEE. Trans. on Power Systems, Vol. 4, pp. 1484-1491, 1989.
- [14] A.D. PAPALEXOPOULOS & C.T. HESTERBERG, - "A regression based approach to short-term system load forecasting", IEEE Trans. on Power Systems, Vol. 5. pp 1535-1547, Nov. 1990.
- [15] G.E.P. BOX & G.M. JENKINS, - "Time Series Modeling Analysis -Forecasting and control", Holden Day, 1970.
- [16] K.W. HIPEL, A.I. McLEOD, - "Time Series Modeling of Water Resources and Enviromental System", Elsevier, 1994.
- [17] M.T. HAGAN & S.M. BEHR, - "The time series approach to short-term load forecasting", IEEE Trans. on Power Systems, Vol.2, No.3, pp. 785-791, August 1987.
- [18] G.LAMBERT TORRES, R. KOSZTALUK & D. MUKHEDKAR, - "Short-term Feeder Load Forecasting using Fuzzy Environment", System Science Journal, Vol. 17. No.2, pp. 51-62, 1991.
- [19] B. VALIQUETTE, G. LAMBERT TORRES & D. MUKHEDKAR, - "An Expert System Based Diagnosis and Advisor Tool for Teaching Power System Operation Emergency Control Strategies", IEEE. Trans. on Power Systems, Vol.6, No.3, pp. 1315-1322, August 1991.
- [20] C.O. TRAORE, G. LAMBERT TORRES & D. MUKHEDKAR, - "Application of Neural Networks for Load Forecasting", Capítulo X, Power Systems for the Year 2000 and Beyond, por M.V. HARIHARAN & J.K. PARIKH, Vedams Books International, July 1991.
- [21] A.P. ALVES DA SILVA, A.M. LEITE DA SILVA, J.C.S. DE SOUZA, M.B. DO COUTO FILHO, - "State Forecasting Based on Artificial Neural Networks", 11th Power Systems Computation Conference (PSCC)", Avignon, France, 1993.
- [22] D. NIEBUR, G. LAMBERT TORRES, A.P. ALVES DA SILVA, et al. - "Artificial Neural Networks for Power Systems - A Literture Survey", International Journal on Engineering Intelligent Systems, Vol.3, No.1, pp. 133-138, December 1993.
- [23] S. RAHMAN & G. SHRESTA, - "A priority vector based technique for load forecasting", IEEE. Trans. on Power Systems, Vol.6, pp. 1459-1465, 1991.
- [24] G. LAMBERT TORRES, C.O. TRAORE, F.G. MANDOLESI, D. MUKHEDKAR, - "Short-term Load Forecasting Using a Fuzzy Engineering Tool", 1st International Forum on Applications of Neural Networks to Power Systems, Seattle, USA, 1991.

- [25] G. LAMBERT TORRES, A.P. ALVES DA SILVA, J. HADDAD, L.O. MATTOS DOS REIS, - "Um Sistema Híbrido para a previsão de carga a curto-prazo", 1º Congresso Brasileiro de Redes Neurais, pp. 307-312, Itajubá, Outubro 1994.
- [26] S.RAHMAN, O. HAZIN, -"A Generalized Knowledge-Based Short-Term Load Forecasting Technique", IEEE. Trans. on Power Systems, Vol.8, No.2, pp. 508-514, May 1993.
- [27] P.C. GUPTA, K. YAMADA, - "Adaptive Short-Term Forecasting of hourly Loads Using Weather Information", IEEE. Trans. Power App. Syst., Vol. PAS-91, No.5, pp. 2085-2094, Sept./Oct. 1972.
- [28] SHUN-ICHI AMARI, - "Mathematical Foundations of Neurocomputing", Proc. IEE, Vol.78, No.9, pp. 1443-1463, Sept. 1990.
- [29] RICHARD P. LIPPMANN, - "An Introduction to Computing with Neural Nets", IEEE ASSP Magazine, pp. 4-22, April 1987.
- [30] YOSHIYASU TAKEFUJI, -"Neural Network Parallel Computing", Kluwer Academic Publishers, 2th edition, pp. 171-176, 1993
- [31] BERNARD WIDROW, & MICHAEL A. LEHR, - "Perceptron, Madaline, and Backpropagation", Proc. IEEE, Vol.78, No.9, pp. 1415-1452, Sept. 1990.
- [32] PAUL J. WERBOS, - "Backpropagation Through Time", Proc. IEEE, Vol.78, No.10, pp. 1550-1560, Oct. 1990.
- [33] TEUVO KOHONEN, - "The Self-Organizing Map", Proc. IEEE, Vol.78, No.9, pp. 1464-1480, September 1990.
- [34] Y.H. PAO, - "Adaptive Pattern Recognition and Neural Network", Addison-Wesley Pub.Co. Inc., Reading, MA., 1989.
- [35] Neuralworks Professional II, User's Guide.
- [36] D.C. PARK, M.A. EL-SHARKAWI, R.J. MARKS II, L. ATLAS & M.J. DAMBORG, - "Electric load forecasting using an artificial neural network", IEEE Trans. on Power Systems, Vol.6, No.1, pp. 442-449, 1991.
- [37] S.J. FARLOW, - "Self-Organizing Methods in Modeling", Marcel Dekker, Inc., 1984
- [38] PENG, N.F. HUBELE & G.G. KARADY, - "Conceptual approach to the application of neural network for short-term load forecasting", Proc. IEEE International Symposium on Circuits and Systems, (ISCAS), New Orleans, LA, USA, 1990, pp. 2942-2945.

- [39] K.Y. LEE, Y.T. CHA & J.H. PARK, - "Short-term load forecasting using an artificial neural network", IEEE Trans. on Power Systems, Vol. 7, No.1, pp. 124-132, February 1992.
- [40] T.M. PENG, N.F. HUBELE & G.G. KARADY, - "Advancement in the application of neural networks for short-term load forecasting", IEEE Trans. on Power Systems, Vol.7, No.1, pp. 250-257, February 1992.
- [41] K.L. HO, Y.Y. HSU & C.C. YANG, - "Short-term load forecasting using a multilayer neural network with an adaptive learning algorithm", IEEE Trans. on Power Systems, Vol.7, No.1, February 1992.
- [42] S.T. CHEN, D.C. YU & A.R. MOGHADDAMJO, - "Weather sensitive short-term load forecasting using non-fully conected artificial neural network", IEEE trans. on Power Systems, Vol.7, No.3, August 1992.
- [43] Y.Y. HSU & C.C. YANG, - "Design of artificial neural networks for short-term load forecasting. Part I: Self-organising feature maps for day type identification", Proc. IEE, Vol.138, No.5, pp. 407-413, September 1991.
- [44] Y.Y. HSU & C.C. YANG, - "Design of artificial neural networks for short-term load forecasting. Part II: Multilayer feddforward networks for peak load and valley load forecasting", Proc. IEE, Vol.138, No.5, pp. 414-418, September 1991.
- [45] A.D. PAPALEXOPOULOS, S. HAO & T.M. PENG, - "An implementation of a neural network based load forecasting model for the EMS", IEEE Trans. on Power Systems, Vol.9, No.4, pp. 1956-1962, November 1994.
- [46] A.G. TEJEDOR, M.J. COSCULLUELA, C. BERMEJO & R. MONTES, - "A neural sistem for short-term load forecasting based on day-type classification", IEEE Trans. on Power Systems, Vol.9, No.4, pp. 353-360, November 1994.
- [47] STATISTIC, Versão 5.0, User's Guide.
- [48] ANDREAS S. WEIGEND & NEIL A. GERSHENFELD, - "Times Series Prediction", Addison Wesley Pub. Co. Inc., Reading, MA., 1992.
- [49] STATISTICA FOR WINDOWS, Versão 4.0, User's Guide.
- [50] L.W. CHAN & F. FALLSIDE, - "An adaptive training algorithm for backpropagation networks", Computer speech and language, Vol.2, pp. 205-218, 1993
- [51] G. LAMBERT TORRES, A. P. Alves da Silva, C. F. COLMÁN & S.M. SOARES. "An hybrid sistem for load forecasting; proposition and tests", Applications of Intelligent software Systems in Power Plant, Process Plant, and Structural Engineering, São Paulo-Brasil, Agosto de 1995.

DATA 05 / 08 / 1996
 PROC. _____
 ED. PPG
 LIV. _____
 RS *Deocá*

I V N O	
BIM - EFEI	
DATA	U I O

FICHA 01 T.841

EFEI / Biblioteca Mauá

Terl 621.316:519.711
C955p

CRÚZ, Carlos Fermín Colmán
Previsão da Carga Elétrica a
Curto-Prazo via Redes Neurais.

Data	N.o Registro	Data
Retirada		Devolução
	96 Adriano Pau	

EFEI

BIBLIOTECA MAUÁ

Esta publicação deverá ser devolvida dentro do prazo estipulado.

O leitor é responsável pela publicação em seu poder.

C

(scribbles)

EFEI - BIBLIOTECA MAUÁ

8200841



NÃO DANIFIQUE ESTA ETIQUETA