

2 Revisão Bibliográfica dos Modelos de Previsão

Neste capítulo, são abordados alguns modelos e conceitos utilizados na literatura para realizar previsão de carga elétrica. Dentre os modelos lineares existentes, serão examinados os modelos ARIMA de Box & Jenkins e amortecimento exponencial de Holt & Winters. Serão estudados também modelos não-lineares provenientes de técnicas de inteligência artificial (Redes Neurais e Lógica Fuzzy). O capítulo se encerra com a descrição da metodologia de combinação de previsões.

2.1. Conceitos Básicos

As informações que produzem uma seqüência de dados distribuídos no tempo são denominadas por séries temporais. Essas podem ser definidas como um conjunto de observações ordenadas no tempo com presença de dependência serial. A forma mais simples de denotar uma série temporal é $y_1, y_2, y_3, \dots, y_t$, que indica uma série de tamanho t . O instante t geralmente indica o último instante disponível.

Seja $Y_{1:t} = (y_1, y_2, y_3, \dots, y_t)$ o conjunto representando t observações passadas da variável aleatória y , conhecida também como série histórica. O objeto de interesse é a determinação das relações de dependência temporal da série através de uma análise estatística de sua série histórica $Y_{1:t}$. Em Fernandes (1995), este estudo é chamado de análise de séries temporais e suas etapas são:

- Investigação do mecanismo gerador da série temporal;
- Descrição do comportamento da série;
- Procura de periodicidades relevantes nos dados;
- Formulação de previsões sobre valores futuros da série.

As três primeiras etapas consistem na determinação das relações de dependência de $Y_{1,t}$ e na elaboração de um modelo que represente a série ajustada. A partir disso, podem-se realizar previsões, por exemplo, a determinação dos futuros valores que assumirão as variáveis aleatórias y_{t+1} , y_{t+2}, \dots, y_{t+k} , sendo k o horizonte de previsão máximo do modelo.

Existem procedimentos variados para estimar um valor futuro, a partir da combinação de valores passados. Conforme descrito por Souza (1993), os modelos de previsão podem ser agrupados em três categorias baseadas no número de séries temporais envolvidas na modelagem:

- Modelos Univariados: baseiam-se somente em uma série temporal;
- Modelos Multivariados: a série de interesse é explicada pelo seu passado e por outras séries temporais não correlacionadas entre si;

Nesta dissertação, serão estudados somente os modelos univariados. Como exemplo, a Figura 2.1 contém um esquema ilustrativo que caracteriza a classe desses modelos.



Figura 2.1- Esquema ilustrativo dos modelos univariados

2.2. Modelos Box & Jenkins

O modelo de previsão de Box et al. (1994) é conhecido como ARIMA (Autoregressive Integrated Moving Average). Esta metodologia, como modelo univariado, tem a finalidade de encontrar uma equação que represente a série temporal y_t , por meio de uma estrutura dependente dos seus valores passados $y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-p}$ e seus erros de previsão um passo à frente, $e_{t-1}, e_{t-2}, e_{t-3}, \dots, e_{t-q}$, onde $e_t = y_t - \hat{y}_{t-1}(1)$. Na sua essência, o modelo assume

que a série temporal é uma realização particular de um processo estocástico, gerado pela passagem sucessiva de um processo ruído branco a_t a uma seqüência de filtros lineares, como mostra a Figura 2.2 a seguir:

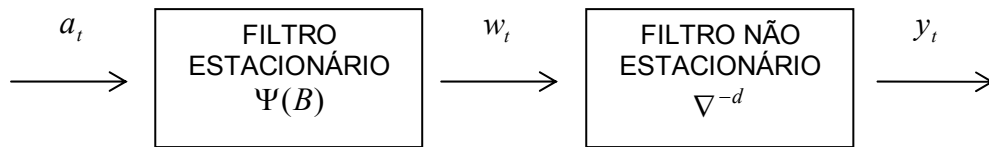


Figura 2.2 - Diagrama do modelo Box & Jenkins

Onde os filtros lineares são definidos por:

$$2.1 \quad \Psi(B) = \theta(B)/\phi(B) \quad \text{Filtro Estacionário}$$

$$2.2 \quad \nabla^{-d} = (1 - B)^{-d} \quad \text{Filtro Não Estacionário}$$

Seja B o operador retardo cujo efeito sobre uma variável discreta é $B^k y_t = y_{t-k}$. O polinômio $\theta(B)$ representa a parte referente às médias móveis do processo a_t , enquanto que o polinômio $\phi(B)$ caracteriza a parte auto-regressiva do modelo. Ambos os polinômios, $\theta(B)$ e $\phi(B)$, são do tipo $P(X) = 1 - C_1 X^1 - C_2 X^2 - \dots - C_k X^k$ com graus q e p, respectivamente. O operador ∇ representa o procedimento diferença tal que $\nabla y_t = y_t - y_{t-1}$. Sendo assim, o modelo ARIMA(p,d,q) é representado pela eq. (2.3).

$$2.3 \quad \phi(B)\nabla^d y_t = \theta(B)a_t$$

Onde d é o grau de diferença da variável y_t .

Em Souza (1996), encontra-se a descrição detalhada dos procedimentos necessários à construção dos modelos ARIMA. Inicialmente, o modelo é identificado a partir da análise das auto-correlações (FAC) e auto-correlações parciais (FACP) dos dados históricos. Posteriormente, os parâmetros

desconhecidos do modelo são estimados para, em seguida, serem testados na adequação do mesmo aos dados. O último passo é realizar as previsões de observações futuras.

Desta forma, o ajuste de um modelo baseado no método Box & Jenkins segue os seguintes passos:

2.2.1. Identificação do Modelo

Segundo Pellegrini (2000), a identificação do modelo mais adequado para descrever a série temporal não é uma tarefa simples, uma vez que existe grande variedade de modelos ARIMA capazes de representar a série. As principais ferramentas usadas no processo de identificação do modelo são a função de auto-correlação (FAC) e a função de auto-correlação parcial (FACP). Na prática, a identificação do melhor modelo consiste teoricamente de quatro estágios:

1. Para algumas séries, é necessário fazer transformações nos dados, com o intuito de torná-las homocedásticas. Um método recomendado é a transformação de Box & Cox, encontrado em Box et al. (1994).
2. Verificar a estacionariedade da série. Este procedimento é feito analisando-se a série temporal e os gráficos das FAC e FACP. Se a série exibir valores em torno de uma constante e os gráficos das FAC e das FACP apresentarem auto-correlações que tendem a zero rapidamente, haverá indícios de que a série seja estacionária. Se algum destes requisitos não for observado, a série é possivelmente não-estacionária.
3. Nesta fase, é feita a escolha adequada do grau de diferenciação d . Séries não-estacionárias devem ser estabilizadas através de diferenciação. Faz-se a diferenciação simples das observações quantas vezes forem necessárias para torná-las uma série estacionária.
4. Para finalizar a etapa da identificação, basta determinar as ordens p e q , das componentes AR e MA do modelo. Para tal, pode-se usar o quadro resumo de Souza (1996), mostrado na Figura 2.3.

Modelo AR(p)
Autocorrelação: decrescente
Autocorrelação Parcial: nula para lags acima de p.

Modelo MA(q)
Autocorrelação: nula para lags acima de q.
Autocorrelação Parcial: decrescente

Modelo ARMA(p,q)
Autocorrelação: é uma mímica do comportamento de AR(p) depois (p-q) lags.
Autocorrelação Parcial: é uma mímica do comportamento de um MA(q) para lags superiores a (p - q).

Figura 2.3 - Comportamento da FAC e FACP nos modelos ARMA

2.2.2. Estimação dos Parâmetros do Modelo

Após a etapa de identificação da ordem do modelo, os parâmetros precisam ser estimados. O método dos mínimos quadrados pode ser usado na identificação dos parâmetros auto-regressivos AR, porém como os parâmetros da média móvel são não-lineares, os processos de estimação se tornam mais complexos. Desta forma, uma outra técnica frequentemente utilizada é o método de máxima verossimilhança.

Essa é uma técnica de estimação bastante usada em inferência estatística, pois os estimadores obtidos apresentam algumas propriedades convenientes. Para grandes amostras, os estimadores de máxima verossimilhança possuem as propriedades assintóticas de:

- Eficiência, isto é, com menor variância em relação a qualquer outro estimador;
- Consistência;
- No caso de haver um estimador suficiente, ele será função do estimador de máxima verossimilhança;
- Ter distribuição aproximadamente normal, com facilidade de cálculo para a obtenção dos parâmetros média e variância.

O princípio da verossimilhança afirma que, se o modelo estiver corretamente identificado e estimado, todas as informações provenientes dos dados observados estão contidas na função de verossimilhança, e todos os demais aspectos informativos das observações são irrelevantes.

Para os leitores que desejam maiores informações sobre as definições da função de máxima verossimilhança e suas propriedades de estimação, recomenda-se a leitura de Bolfarine (2001).

2.2.3. Validação do Modelo

Depois da estimação dos parâmetros do modelo, devem-se aplicar testes estatísticos para comprovar sua validade. Um método inicial para verificar o modelo ajustado é analisar os resíduos do modelo, representados pelos erros de previsão um passo à frente, $e_t = y_t - \hat{y}_{t-1}(1)$, e a função de auto-correlação dos mesmos (Montgomery, 1976). Segundo (Souza 1996), se o modelo obtido for apropriado, a FAC da amostra dos resíduos $\hat{\rho}(e_t)$ não deve ser significativa, ou seja, não deve apresentar valores superiores a $n^{-1/2}$ para nenhum período. Se este for o caso, os valores observados na série temporal y_t foram transformados em um processo de ruído aleatório e_t .

Um segundo método de validação do modelo é o teste de *Sobrefixação*, que consiste na construção de um modelo com um número de parâmetros superior ao do modelo fixado. Este modelo mais elaborado é submetido à estimação dos seus parâmetros, que demonstrará a necessidade ou não de parâmetros adicionais. Desta forma, tenta-se corrigir algum erro que tenha sido cometido na etapa de identificação do modelo.

Uma outra forma de verificar a significância dos modelos é submeter a testes os resíduos como, por exemplo, os testes de *Portmanteau*. Considerando os parâmetros estimados do modelo $\hat{\theta}(B)$ e $\hat{\phi}(B)$ para um ARIMA (p,d,q), pode-se estimar os resíduos através da expressão:

$$2.4 \quad \hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)w_t$$

Onde a variável w_t substitui $\nabla^d y_t$. Se o modelo fixado for adequado, então os resíduos estimados \hat{a}_t tendem para uma série ruído branco de média nula e variância $\hat{\sigma}_a^2$.

Conforme já mencionado, a partir das séries dos resíduos estimados podem-se calcular as k primeiras auto-correlações dos resíduos, isto é:

$$\hat{\rho}_i(\hat{a}); i = 0, 1, 2, \dots, k$$

Os testes *Portmanteau* não consideram os valores de $\hat{\rho}_i(\hat{a})$ isoladamente, mas o conjunto dos k primeiros $\hat{\rho}_k(\hat{a})$. O teste sugerido por Box & Jenkins (Souza, 1996), assume que, se o modelo fixado for correto, então a estatística:

$$2.5 \quad Q = n \sum_{i=1}^k \hat{\rho}_i^2(\hat{a})$$

segue aproximadamente uma distribuição Qui-Quadrado (χ^2), e testa-se a hipótese de um conjunto de auto-correlações residuais ser significativo.

Existem ainda outros testes que podem ser feitos, como por exemplo, a análise do periodograma acumulado da série dos resíduos estimados \hat{a}_t , para verificar a existência de componentes periódicos na série dos \hat{a}_t .

2.2.4. Previsão

Após o término do processo de identificação, estimação dos parâmetros e verificação, o último passo é realizar as previsões de observações futuras. Partindo-se da origem em t , e supondo que o objetivo é prever a série em um período futuro k , $\hat{y}_t(k)$ representa a previsão para um período $t+k$ feita em t .

A previsão múltiplos passos para o período $t+k$ é normalmente construída a partir de sucessivas previsões para os períodos $t+1, t+2, \dots, t+k-1$ (Montgomery, 1976). Neste procedimento, o valor de y_{t+k} , o qual não se conhece no tempo t , é substituído pela sua previsão $\hat{y}_t(k)$. Portanto, a previsão se torna o cálculo do

valor esperado de uma futura observação condicionada aos valores passados e ao valor presente da variável. Ou seja, $\hat{y}_t(k)$ é o valor previsto para um horizonte k períodos de tempo futuro e t o período de origem da previsão, então,

$$2.6 \quad \hat{y}_t(k) = E(y_{t+k} | y_t, y_{t-1}, \dots)$$

Como o valor de a_{t+k} na equação de previsão não é conhecido no tempo t , deve-se substituir por zero, e os valores passados por $a_{t-j} = y_{t-j} - \hat{y}_{t-j-1}(1)$.

2.2.5. Modelos SARIMA

Os modelos ARIMA são formados a partir da correlação entre os valores de y_t observados em instantes de tempo consecutivos (Sartoris, 2000). Porém, quando a periodicidade da série temporal for inferior a um ano, outro tipo de correlação serial passa ser significativa: a correlação entre instantes de tempo distantes entre si por s ou múltiplos de s , onde s é o número de observações contidas em um ano. Por exemplo, para dados mensais deve-se considerar s igual a 12, enquanto que se a série for semanal, s passa a ser 53.

Para acomodar esse tipo de série, surgem os modelos ARIMA sazonais, conhecidos também como SARIMA (*Sazonal Autoregressive Integrated Moving Average*). Neste caso, a modelagem segue a eq. 2.7:

$$2.7 \quad \phi(B)\Phi(B^s)\nabla_s^D\nabla^d y_t = \theta(B)\Theta(B^s)a_t$$

Onde o polinômio $\Phi(B^s)$ representa o operador sazonal auto-regressivo com Φ_i parâmetros auto-regressivos sazonais e i igual a 1, 2, ..., P. O operador sazonal de médias móveis $\Theta(B^s)$ possui Θ_j parâmetros de médias móveis sazonais, com j possuindo valor máximo Q.

O operador diferença sazonal ∇_s^D altera a variável discreta y_t de forma semelhante ao operador diferença não-sazonal. A variável é diferenciada pelo valor s passos atrás, isto é, $\nabla_s^D = (1 - B^s)^D$.

Os instrumentos utilizados para a identificação de modelos sazonais continuam sendo a FAC e FACP. O procedimento de obtenção deste modelo segue os mesmos passos empregados para determinar o modelo ARIMA. Isto é, na modelagem SARIMA, faz-se também a análise do comportamento da FAC e FACP, com a ressalva de auto-correlações significantes nas defasagens sazonais. Como por exemplo, as séries mensais devem possuir auto-correlações em 12, 24, 36, etc.

As demais etapas da metodologia de Box & Jenkins, aplicadas a modelos sazonais, são análogas àquelas descritas para os modelos não-sazonais.

2.3. Modelos de Amortecimento Exponencial

Esteves (2003) relata que os modelos de amortecimento exponencial são utilizados largamente para previsão de séries temporais, devido à sua simplicidade, facilidade de implementação computacional e bons resultados. São métodos que usam uma ponderação diferente em cada valor observado da série temporal, de modo que os valores mais recentes recebam pesos maiores. Assim, os pesos atribuídos às observações decaem exponencialmente a partir de valores mais recentes.

Os modelos de Holt & Winters (Souza, 1983) e (Montgomery, 1976) são apropriados para séries temporais com o sazonalidade. Dependendo da natureza da série temporal, podem-se usar duas formulações diferentes: forma aditiva ou multiplicativa.

Nos dois tipos de modelo que serão apresentados, os parâmetros serão definidos por:

b_1 = Componente do Nível, também conhecido como Componente Constante;

b_2 = Componente de Tendência Linear

c_t = Componente Sazonal

ε_t = Componente de erro aleatório (Ruído)

2.3.1. Modelo Sazonal Multiplicativo

O modelo multiplicativo é utilizado em dados sazonais nos quais a amplitude do ciclo sazonal varia com o passar do tempo. Sua representação é expressa por:

$$2.8 \quad y_t = (b_1 + b_2 t)c_t + \varepsilon_t$$

O comprimento do ciclo sazonal é de S períodos e as componentes sazonais são definidas de maneira que a soma das mesmas resulte em S , ou seja,

$$2.9 \quad \sum_{t=1}^S c_t = S$$

A Figura 2.4 mostra uma série temporal para a qual um modelo multiplicativo é apropriado. Percebe-se que enquanto a tendência da série $(b_1 + b_2 t)$ aumenta, a amplitude da sazonalidade também aumenta.

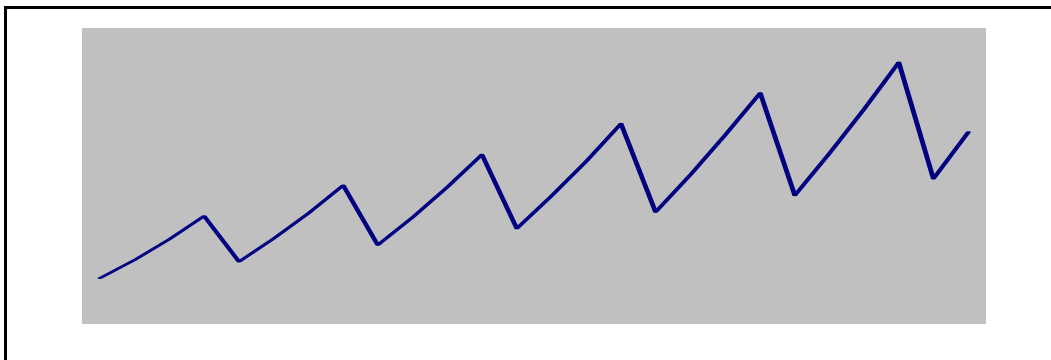


Figura 2.4 - Série típica com aumento de padrão sazonal

Definem-se os valores estimados para a tendência linear e a componente sazonal no final do período t , como $\hat{b}_2(t)$ e $\hat{c}_t(t)$ respectivamente. A componente constante (ou de nível) $\hat{b}_1(t)$ é estabelecida na origem corrente t e estimada por $\hat{a}_1(t)$.

O procedimento contínuo de atualização dos parâmetros do modelo e de previsão é relativamente simples. Ao final de cada período t , depois da realização da observação y_t , os parâmetros são calculados pelas seguintes expressões:

Atualização da componente constante

$$2.10 \quad \hat{a}_1(t) = \alpha \frac{y_t}{\hat{c}_t(t-S)} + (1-\alpha)[\hat{a}_1(t-1) + \hat{b}_2(t-1)]$$

Atualização da componente de tendência linear

$$2.11 \quad \hat{b}_2(t) = \beta[\hat{a}_1(t) - \hat{a}_1(t-1)] + (1-\beta)\hat{b}_2(t-1)$$

Atualização da componente sazonal

$$2.12 \quad \hat{c}_t(t) = \gamma \frac{y_t}{\hat{a}_1(t)} + (1-\gamma)\hat{c}_t(t-S)$$

Os modelos de Holt–Winters são caracterizados por três constantes de amortecimento: α , β e γ , definidos como hiperparâmetros utilizados na atualização dos parâmetros $\hat{b}_1(t)$, $\hat{b}_2(t)$ e $\hat{c}_t(t)$ respectivamente.

Para realizar a previsão da variável estudada no período $t + k$,

$$2.13 \quad \hat{y}_{t+k}(t) = (\hat{b}_1(t) + k\hat{b}_2(t))\hat{c}_{t+k}(t+k-S)$$

2.3.2. Modelo Sazonal Aditivo

Já o modelo aditivo é utilizado na modelagem de dados sazonais nos quais a amplitude do ciclo sazonal permanece constante com o passar do tempo. Sua equação é:

$$2.14 \quad y_t = b_1 + b_2 t + c_t + \varepsilon_t$$

Como no modelo multiplicativo, o comprimento do ciclo sazonal também será definido por S períodos. Porém, este modelo é apropriado para séries temporais nas quais a amplitude do ciclo sazonal é independente da tendência da série ($b_1 + b_2 t$). A partir desta característica, (Montgomery, 1976) define a soma das componentes sazonais igual a zero, ou seja,

$$2.15 \quad \sum_{t=1}^S c_t = 0$$

A Figura 2.5 mostra uma série temporal para um modelo sazonal aditivo. Percebe-se que, apesar da tendência crescente da série, a variação da sazonalidade se mantém constante.

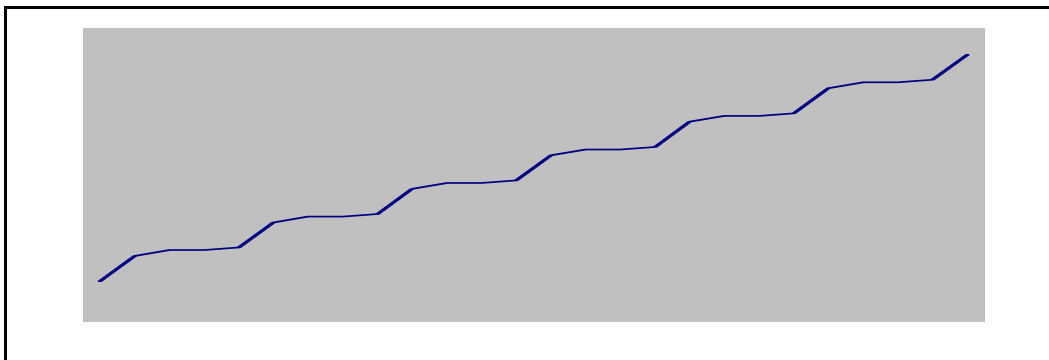


Figura 2.5 - Série típica para o modelo aditivo

Considere-se que os valores estimados para a tendência linear e a componente sazonal no final do período t são definidos por $\hat{b}_2(t)$ e $\hat{c}_t(t)$ respectivamente. A componente constante é estabelecida na origem corrente t e estimada por $\hat{a}_1(t)$.

Os procedimentos de atualização dos parâmetros do modelo e a equação de previsão são expressos pelas seguintes equações:

Atualização da componente constante

$$2.16 \quad \hat{a}_1(t) = \alpha[y_t - \hat{c}_t(t - S)] + (1 - \alpha)[\hat{a}_1(t - 1) + \hat{b}_2(t - 1)]$$

Atualização da componente de tendência linear

$$2.17 \quad \hat{b}_2(t) = \beta[\hat{a}_1(t) - \hat{a}_1(t-1)] + (1 - \beta)\hat{b}_2(t-1)$$

Atualização da componente sazonal

$$2.18 \quad \hat{c}_t(t) = \gamma[y_t - \hat{a}_1(t)] + (1 - \gamma)\hat{c}_t(t - S)$$

A equação da tendência é a mesma utilizada para o modelo multiplicativo da eq. (2.11). Já as equações de atualização das componentes do nível e do fator sazonal são diferentes. A diferença é que a componente sazonal está efetuando operações de soma e subtração, ao invés de multiplicar e dividir. Para realizar a previsão da variável estudada no período $t + k$, emprega-se a equação:

$$2.19 \quad \hat{y}_{t+k}(t) = \hat{b}_1(t) + \hat{b}_2(t)k + \hat{c}_{t+k}(t + k - S)$$

2.4. Modelos de Redes Neurais

As redes neurais artificiais (RNA's) consistem em um método no qual constrói-se um sistema que simula o funcionamento do cérebro humano e busca reconhecer regularidades e padrões nos dados que lhe são apresentados (Fernandes, 1995). Devido à sua capacidade em manipular dados imprecisos e ruidosos, as RNA's vêm se tornando uma poderosa ferramenta para modelagem de séries temporais de carga elétrica. Muitas aplicações têm sido desenvolvidas nos últimos anos nesta área, e a maioria destas aplicações tem mostrado bons resultados quando comparadas com outros procedimentos estatísticos. Uma abordagem completa pode ser encontrada em Haykin (1999). Alguns exemplos desta modelagem podem ser encontrados nos trabalhos expostos na Tabela 2.1, que modelam séries temporais horárias, diárias e mensais.

Autor	Título
(Hippert 2001)	Previsão de Cargas a Curto Prazo
(Hippert 2000)	Previsão de Temperatura Horárias
(Serão 2003)	Previsão de Carga de Curto Prazo
(Sobral 2003)	Previsão de Carga de Curto Prazo
(Rizzo 2001)	Previsão de Carga de Curtíssimo Prazo
(Fonte 2002)	Previsão de Elevação de Temperatura
(Pinto 2002)	Previsão de Séries Temporais Lineares
(Teixeira 1999)	Previsão de Carga

Tabela 2.1 - Trabalhos publicados em previsão de séries temporais

2.4.1. Componentes da RNA

As redes são compostas por unidades de processamento (neurônios artificiais), que se interconectam, formando redes capazes de armazenar e transmitir informações provenientes do meio externo. Uma RNA é caracterizada por três tópicos básicos:

- Neurônio Artificial: Componentes básicos, inspirados no neurônio biológico;
- Topologia: Maneira de conexão entre os diversos neurônios artificiais que formam a RNA;
- Treinamento: É o processo pelo qual a RNA aprende, por meio da apresentação dos padrões desejados.

2.4.1.1. Neurônio Artificial

Trata-se de um modelo correspondente à estrutura simplificada do neurônio biológico. É constituído por três partes fundamentais: pesos sinápticos, que ponderam os valores das entradas do neurônio; regra de propagação, que define como serão combinadas as entradas do neurônio; e a função de ativação, que determina o efeito que o resultado da regra de propagação terá sobre o nível de ativação do neurônio.

Um exemplo de neurônio artificial e seus componentes podem ser vistos na Figura 2.6.

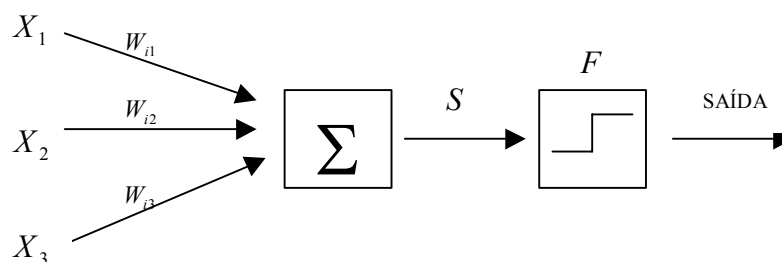


Figura 2.6 - Elemento Processador

Este neurônio foi elaborado para reproduzir algumas características de um neurônio biológico. Por exemplo, as ligações entre dois neurônios vizinhos (sinapses) podem ser representadas por um conjunto X_j de entradas aplicado ao neurônio i , cada uma representando a saída de outros neurônios artificiais. Cada entrada é multiplicada por um peso correspondente W_{ij} , resultando em entradas ponderadas. Em seguida, somam-se todas as entradas, obtendo-se um valor S que representa o único argumento da função F de transferência.

As funções de ativação processam a soma S dando origem à saída final do neurônio. Dentre algumas formas de função de ativação na literatura (Haykin, 1999), pode-se citar, por exemplo, uma função limiar e uma função logística.

A função da Figura 2.7 representa uma função de limiar cuja saída do neurônio possui apenas dois valores: zero para qualquer argumento S menor que zero e um se o argumento for maior ou igual a zero.

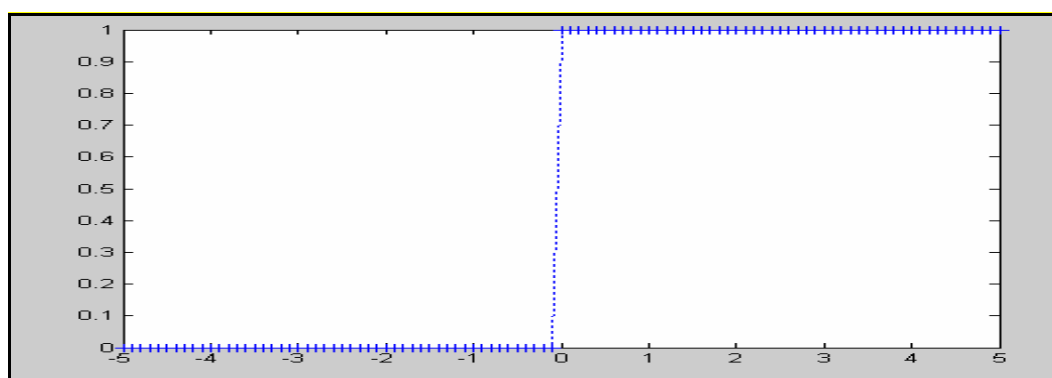


Figura 2.7 - Função de ativação limiar

Lehmann (2002) informa que nos casos onde se deseja aproximar com maior precisão os neurônios biológicos, usam-se funções não-lineares. Um

exemplo desta função é a sigmóide, cuja representação matemática é dada pela expressão:

$$2.20 \quad f(x) = 1/(1 + e^{-x})$$

Esta função é mostrada na Figura 2.8 e possui o domínio $(-\infty; +\infty)$ e imagem entre zero e um.

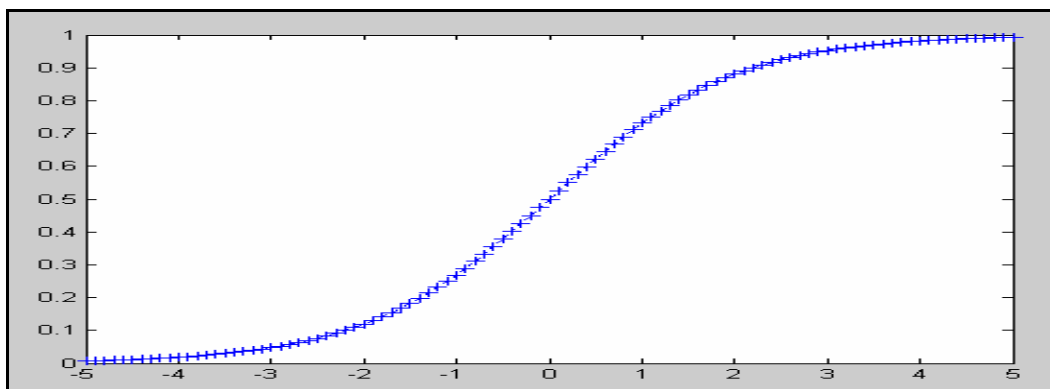


Figura 2.8 - Função de ativação sigmóide

2.4.1.2. Topologia da RNA

A estrutura de organização dos neurônios chama-se de *topologia da RNA*. Cada disposição vertical de neurônios é definida como camada e as redes que possuem mais de uma camada de neurônios são conhecidas como redes Perceptron Múltiplas Camadas, em inglês Multi Layer Perceptron (MLP). Em geral, elas possuem três camadas, sendo a primeira nomeada de camada de entrada, a camada seguinte de camada escondida, e a última recebe o nome de camada de saída.

Basicamente, existem duas classes de topologias: as redes neurais *Não-Recorrentes* e as *Recorrentes*. As primeiras não possuem realimentação de suas saídas para as entradas, enquanto que as recorrentes contêm realimentação das saídas dos neurônios para as entradas, sendo suas saídas determinadas pelas entradas mais atuais e pelas saídas anteriores (Abelém, 1994).

Como exemplo, a Figura 2.9 apresenta uma RNA não-recorrente com a seguinte topologia (5-3-1). Ou seja, a rede possui a camada de entrada com

cinco neurônios, uma camada escondida com três neurônios e a camada de saída com apenas um neurônio.

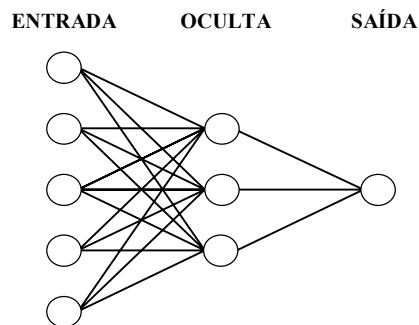


Figura 2.9 - Topologia de uma RNA com uma Camada Escondida

2.4.1.3. Treinamento da RNA

A forma com o qual os pesos W_{ij} são ajustados é chamado de *treinamento da RNA*. Dentre os tipos existentes, destacam-se dois: treinamento supervisionado e não supervisionado.

O treinamento supervisionado baseia-se na apresentação de exemplos conhecidos, isto é, trabalha com dados de entrada e valores desejados de saída simultaneamente. Dos algoritmos que estabelecem, como as conexões são atualizadas, um dos mais conhecidos chama-se *back-propagation*.

Já nas RNA's treinadas de forma não-supervisionada, somente os padrões de entrada estão disponíveis para a rede. Este tipo de aprendizagem se torna possível quando existe abundância nos dados de entrada. Esse tipo de rede neural artificial não requer valores de saída, um exemplo são os mapas auto-organizáveis (Haykin, 1999).

Nesta dissertação, serão utilizadas redes com treinamento supervisionado por *back-propagation*. Esse consiste em duas fases complementares: *Forward* e *Backward*:

- *Forward*: Esta etapa é utilizada para definir a saída da rede para um dado padrão de entrada. Calcula-se o erro entre o valor produzido pela camada de saída e o valor desejado.

- *Backward*: Depois do processamento *Forward*, o erro é propagado a partir da camada de saída até a camada de entrada e os pesos das conexões das unidades das camadas internas vão sendo modificados.

Como apresentado por Martins (2002), o algoritmo *back-propagation* da Figura 2.10 é baseado na regra delta generalizada. Esta regra funciona quando são utilizados neurônios que possuem funções de ativação diferenciável e não-decrescente. Nestes casos, uma função amplamente utilizada é a função sigmóide mostrada na Figura 2.8.

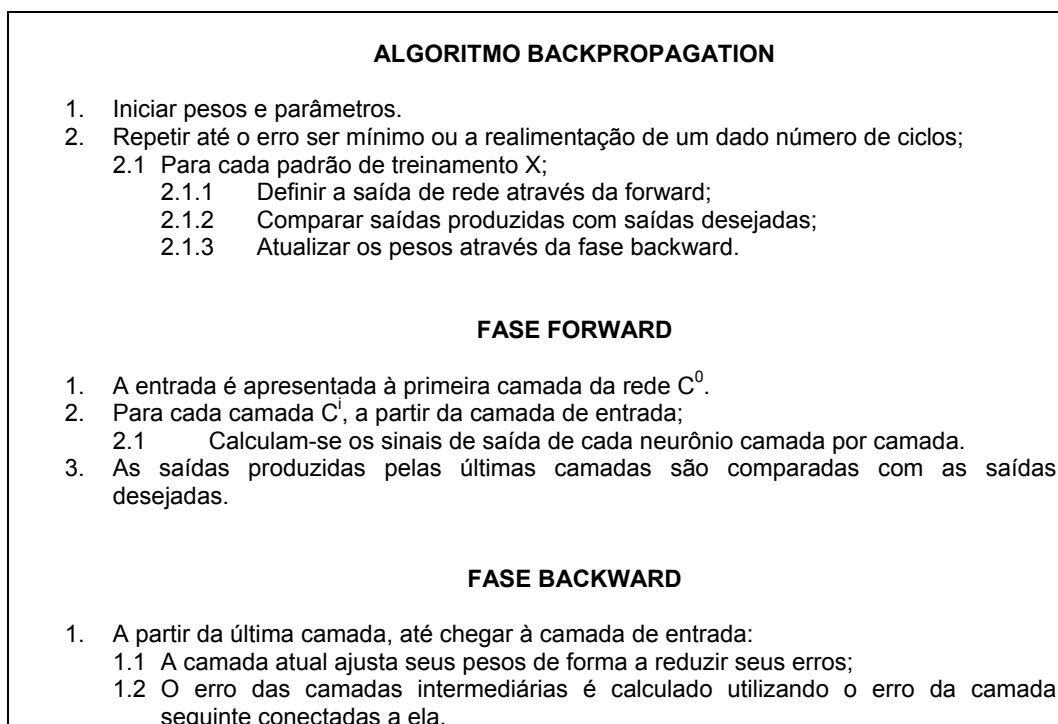


Figura 2.10 - Algoritmo *Backpropagation*

Os ajustes dos pesos são realizados utilizando o método do gradiente descendente. Porém, quando a rede demanda um tempo de treinamento consideravelmente longo, devido ao elevado número de passos no treinamento, podem-se utilizar também os efeitos de segunda ordem para o gradiente decrescente para diminuir o esforço computacional que essa rede demanda.

Depois que a rede estiver treinada e com um nível satisfatório de erro, poderá realizar classificação com novos dados. Porém, isso ocorrerá somente se a rede for utilizada no modo progressivo, ou seja, uma rede não-recorrente. Assim, novas entradas são apresentadas à camada de entrada, processadas nas camadas ocultas e os resultados apresentados na camada de saída, como

no treinamento, mas sem a retro-propagação do erro. A saída apresentada é o modelo dos dados, na interpretação da rede.

Para melhorar o desempenho do treinamento, pode-se utilizar uma constante de momento. Com a inclusão deste termo na equação de ajuste dos pesos sinápticos, a velocidade de aprendizado aumenta substancialmente, diminuindo as chances de instabilidade da rede, além de acelerar o treinamento em regiões muito planas. A utilização da constante de momento possui o benefício de evitar que o processo de aprendizagem termine em um mínimo local na superfície de erro (pontos na superfície de erro que apresentam uma solução estável, embora não sejam saídas desejadas).

2.4.1.4. Previsão de Séries Temporais

Algumas informações provenientes do ambiente externo são de natureza dinâmica e seqüencial, de modo que as relações temporais entre padrões consecutivos precisam ser levadas em consideração. O reconhecimento de tais padrões dinâmicos é crucial para a previsão de séries temporais. Assim, um objeto de estudo na teoria da RNA tem sido os diversos aspectos da aprendizagem serial ou aprendizagem de seqüências temporais (Barreto, 2000).

Para que a RNA seja usada como procedimento de previsão de séries temporais, torna-se necessário à incorporação do tempo para capacitá-la a modelar as variações estatísticas de uma seqüência temporal de dados. Segundo Haykin (1999), esse problema é resolvido através da incorporação de um mecanismo chamado *memória de curto prazo* (MCP). Os componentes individuais da série temporal são armazenados de maneira que o processo de aprendizagem da rede trabalhe em dois eixos: tanto os itens em si, quanto a ordem em que eles ocorreram.

O principal objetivo da memória é transformar uma rede estática em uma rede dinâmica. Com isso, as RNA's com características puramente estáticas se tornam dinâmicas. Ocorre, portanto uma separação de responsabilidades na qual a rede estática é responsável pela não-linearidade e a memória é responsável pelo tempo. A forma de MCP mais simples utilizada é chamada de memória de linha de atraso.

A rede MLP (aprendizado *back-propagation*), com uma memória de linha de atraso derivada, aplicada à entrada, é a rede mais simples e uma das mais

usadas em previsão de séries temporais. A Figura 2.11 apresenta uma estrutura com uma rede MLP e uma janela de tempo.

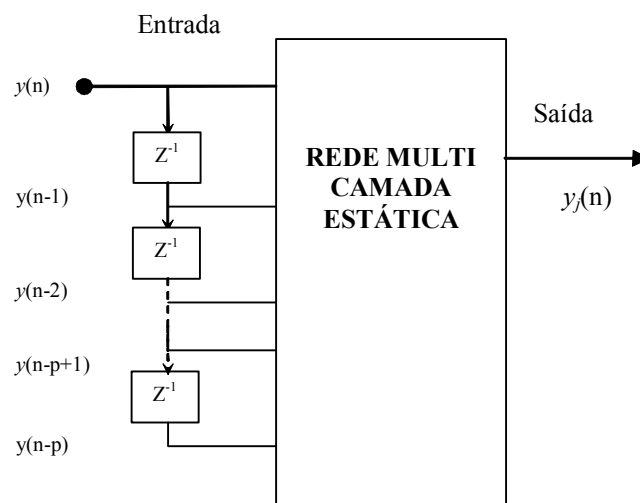


Figura 2.11- Rede MLP com memória de linha de atraso

Cada um dos p blocos constitui, neste caso, um operador de unidade de atraso Z^{-1} , sendo p a ordem da memória de valor igual ao número total de atrasos.

2.5. Modelos de Lógica Fuzzy

Antes de descrever os modelos de previsão utilizando a Lógica Fuzzy (LF), é preciso compreender algumas definições sobre esta teoria. No dicionário Webster's, a palavra fuzzy é traduzida como algo vago ou indistinto, isto é, o antônimo de preciso e exato. Já entre diversos trabalhos publicados na literatura, em (Tanscheit, 2003) encontram-se as seguintes definições:

“A Lógica Fuzzy procura modelar os modos imprecisos do raciocínio que têm um papel fundamental na habilidade humana de tomar decisões”.

“Esta teoria se mostra capaz de capturar informações imprecisas, descritas em linguagem natural e convertê-las para um formato numérico”.

A seguir são apresentadas uma breve descrição da teoria dos Sistemas Fuzzy (SF) e a metodologia utilizada para a aplicação de previsão em séries temporais.

2.5.1. Sistemas Fuzzy

O SF caracteriza-se por possuir uma coleção de variáveis de entrada, uma coleção de variáveis de saída, conjuntos fuzzy associados a cada variável, uma coleção de regras que associa as entradas para resultar em conjuntos para a saída e uma função que desfuzzifique a saída. Para construção do sistema fuzzy são considerados os componentes representados na Figura 2.12.

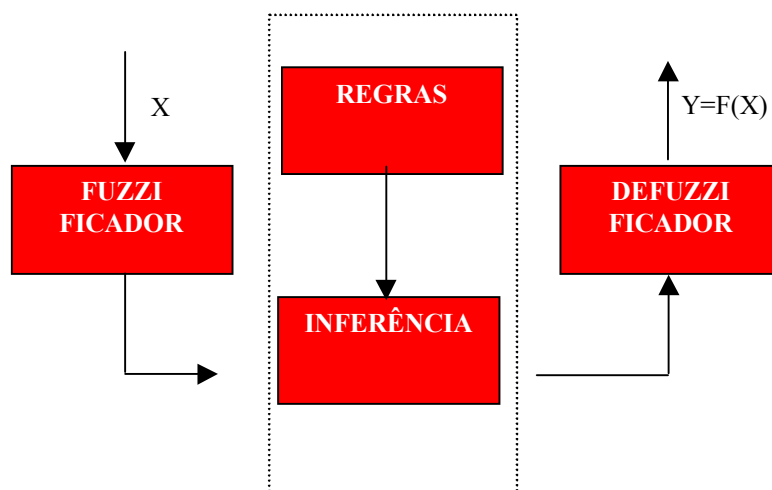


Figura 2.12 - Diagrama de Sistema Fuzzy

Onde:

FUZZIFICADOR – Serve para mapear valores numéricos em conjuntos fuzzy, além de ativar as regras que estão relacionadas às variáveis lingüísticas.

REGRAS – Fornecidas por especialistas ou extraídas de dados numéricos. É formado por um conjunto de implicações do tipo SE - ENTÃO.

INFERÊNCIA – Mapeiam-se conjuntos fuzzy de entrada em um conjunto fuzzy de saída. Determina como as regras são ativadas e combinadas.

DEFUZZIFICADOR – Mapeia o conjunto fuzzy resultante da inferência em um valor numérico preciso. Ou seja, a partir dos graus de participação de cada variável de uma regra, resulta no grau de participação da saída e conseqüentemente no valor real da saída.


A maior vantagem nesta teoria surge no momento em que as variáveis não são tratadas apenas em um estado, mas sim em n estados, cada um com um grau de associação. Formalmente, na teoria clássica dos conjuntos, um elemento pode pertencer ou não a determinado conjunto. Existem três formas de se definir os elementos de um conjunto (Ribeiro, 2002):

1. Pela enumeração de seus elementos $A = \{u_1, u_2, \dots, u_n\}$
2. Por uma propriedade característica $A = \{u / p(u)\}$
3. Por uma função característica $\chi_A : U \rightarrow \{0,1\}$, isto é, $\chi_A(u) = \begin{cases} 1, & \text{se } u \in A \\ 0, & \text{se } u \notin A \end{cases}$


E na teoria do SF, um conjunto fuzzy A do universo de discurso Ω é definido por uma função de pertinência $\mu_A : \Omega \rightarrow [0,1]$. Essa função associa a cada elemento x de Ω o grau $\mu_A(x)$, com o qual x pertence ao conjunto A . A função de pertinência $\mu_A(x)$ indica o grau de compatibilidade do elemento x em relação ao conjunto A . A teoria dos conjuntos fuzzy e a lógica fuzzy constituem a base para a formulação dos Sistemas de Inferência Fuzzy (Mendel, 1992b).

Um outro componente visto na Figura 2.12, é o banco de regras. Estas são definidas por implicações do tipo SE - ENTÃO e envolvem variáveis lingüísticas às quais são atribuídos conjuntos fuzzy. As variáveis lingüísticas de cada regra são agregadas, utilizando conectores lógicos do tipo (E/OU). Um exemplo apresentado em (Neto, 1999) mostra a composição de uma regra:

Se u_1 é quente E u_2 é muito baixo Então v gira um pouco para direita



antecedente



conseqüente

Neste exemplo, u_1 , u_2 e v são variáveis lingüísticas que, recebem como valores, os conjuntos fuzzy “quente”, “muito baixo” e “pouco para direita”,

respectivamente. O resultado de cada regra ativada por um sistema fuzzy é um conjunto fuzzy de saída modificado pelo grau de disparo da regra. A modificação do conjunto fuzzy de saída é determinada de acordo com o método de inferência utilizado. A Figura 2.13 mostra os passos necessários no sistema de inferência fuzzy.

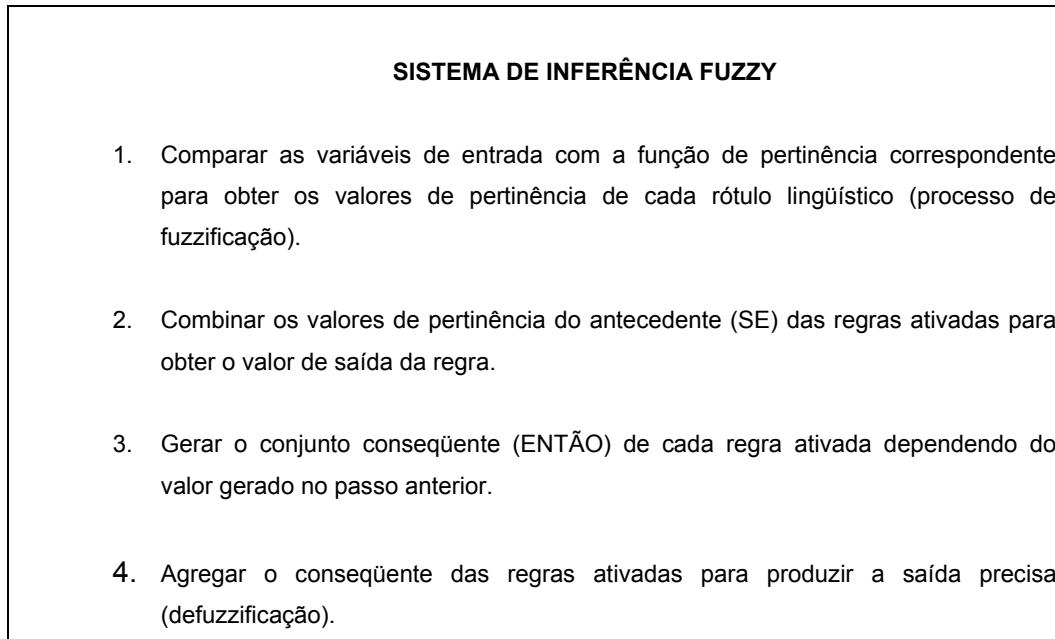


Figura 2.13 - Passos para o processo de inferência

Dos modelos de inferência existentes na construção do SF, pode-se citar:

1. Mamdani (1975), onde as regras são do tipo SE x é A E y é B ENTÃO z é C , sendo A , B , e C conjuntos fuzzy.
2. Sugeno (1985), onde as regras são do tipo SE x é A E y é B ENTÃO z é $f(x,y)$, sendo o conseqüente da regra uma função não fuzzy das variáveis de entrada, em geral um polinômio.

A diferença entre os dois tipos anteriores de inferência fuzzy está no conseqüente das regras. Cada tipo apresenta uma forma de agregação dos resultados de cada regra e métodos de defuzzificação próprio. O modelo de MAMDANI, sendo utilizado nesta dissertação na previsão de carga.

Em um sistema de inferência fuzzy, as regras podem ser formuladas a partir de dados numéricos de entrada e saída de um sistema ou a partir do conhecimento e experiência de um especialista.

2.5.2. Algoritmo de Regras Automáticas de Mendel

Para aplicações em previsão de séries temporais, as regras são estabelecidas pela série histórica analisada, determinando os conjuntos fuzzy do antecedente e do conseqüente referentes às entradas e saídas respectivamente. Uma outra maneira de estabelecer essas regras seria especificar os conjuntos e associar os dados a esses conjuntos definidos (Mendel, 1992b). Da mesma forma que é aplicada na RNA, o procedimento de janela temporal também é utilizado nos SF.

Seja uma seqüência de observações $y(k)$ $k = 1, 2, 3, \dots$, caracterizando uma série temporal. O objetivo da previsão pode ser resumido da seguinte forma: uma vez conhecidos os valores $y(k-n+1), y(k-n+2), \dots, y(k)$, deseja-se encontrar o valor estimado da variável l períodos à frente da origem k , isto é, determinar $y(k+l)$, onde n e l são, respectivamente, a janela temporal e o horizonte de previsão. Assim, deseja-se determinar o seguinte mapeamento:

$$f : (y_{k-n+1}, \dots, y_{k-1}, y_k) \rightarrow y_{k+l}.$$

A Figura 2.14 representa um exemplo de previsão dada uma janela com quatro medidas da variável. A partir de $[y(k-3), y(k-2), y(k-1)$ e $y(k)]$, determina-se o valor para um horizonte, $y(k+1)$.

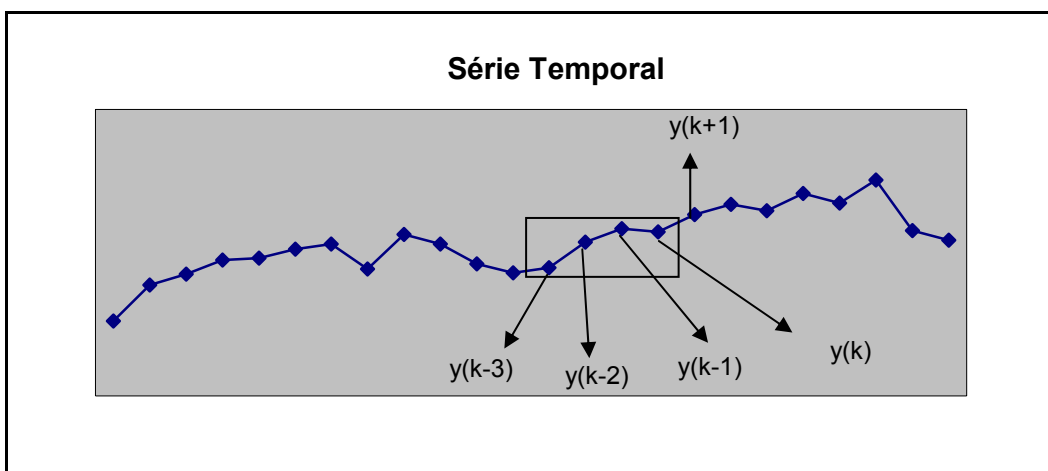


Figura 2.14 - Previsão de Séries Temporais na lógica fuzzy

O algoritmo proposto por Mendel (1992a) consiste em cinco etapas descritas a seguir:

1. Dividir a faixa de valores possíveis da série $[U^-, U^+]$ em m conjuntos fuzzy;
2. Gerar as regras fuzzy, a partir dos pares de dados de entrada e saída coletados. Consiste em cinco passos para cada regra j :
 - a. Determinar o tamanho da janela de tempo n ;
 - b. Determinar o horizonte de previsão l ;
 - c. Determinar o grau de pertinência de cada elemento dos pares de dados de entrada e saída;
 - d. Atribuir a cada elemento dos pares de dados de entrada e saída o conjunto fuzzy com maior grau de pertinência;
 - e. Obter uma regra para par de dados de entrada e saída;
3. Associar um grau para cada regra, multiplicando-se o grau de pertinência de cada termo do antecedente e conseqüente. Caso haja regras conflitantes, isto é, regras compostas dos mesmos antecedentes, porém com conseqüentes distintos, utilizar aquela que possuir maior grau dentre as conflitantes.
4. Criar um banco combinado de regras fuzzy;
5. Determinar o mapeamento baseado no banco combinado de regras fuzzy. Um dos métodos para gerar a saída precisa é o centróide (centro de gravidade).

2.6. Combinação de Previsões

Segundo Granger (1980), se duas ou mais previsões fossem combinadas, o resultado final geraria melhores previsões que cada um dos métodos separadamente. Entretanto, para que a combinação de resultados de previsão seja satisfatória, cada modelo de previsão deve ser não-viesado, no sentido de não subestimar ou sobreestimar consistentemente os valores verdadeiros.

Para entender a teoria da combinação de previsões, supõe-se duas previsões f_i e g_i , ambas um passo à frente para o valor x_{i+i} , que geram erros de previsão ef_i e eg_i .

Além disso, suponha que as previsões os erros ef_i e eg_i possuem média zero e variância constante. Uma previsão combinada dada por

$$2.21 \quad c_i = kf_i + (1-k)g_i$$

Onde $0 < k < 1$, c_i será necessariamente não-viesado e o erro resultante da previsão combinada dado por:

$$2.22 \quad ec_i = x_{i+i} - c_i = kef_i + (1 - k)eg_i$$

O valor k pode ser escolhido tal que a variância dos erros de previsão combinadas, ec_i , seja menor do que qualquer uma das variâncias dos erros dos modelos individuais. Como a média dos erros é nula, conseqüentemente a variâncias e os erros quadráticos médio serão idênticos. A variância de ec_i é dada por:

$$2.23 \quad Var(ec) = k^2Var(ef) + (1 - k)^2Var(eg) + 2k(1 - k)Cov(ef; eg)$$

onde a $Cov(ef; eg)$ é a covariância entre os erros de f_i e g_i . Para minimizar a $Var(ec)$, basta derivar a eq. 2.23 com relação a k e igualar a zero:

$$2.24 \quad k = \frac{Var(eg_i) - Cov(ef_i; eg_i)}{Var(eg_i) + Var(ef_i) - 2Cov(ef_i; eg_i)}$$

Se ef_i e eg_i forem descorrelatados, essa equação se reduz a:

$$2.25 \quad k = \frac{Var(eg_i)}{Var(eg_i) + Var(ef_i)}$$

Se substituirmos a eq. 2.25 na equação da variância dos erros de previsão combinada (eq. 2.23), concluí-se que a variância resultante para os erros de previsão combinadas será menor do que ambas as variâncias de ef_i e eg_i , logo a variância nunca será maior do que a menor das variâncias das duas previsões.

Assim, em teoria, uma previsão combinada será geralmente superior a qualquer dos métodos de previsão e nunca poderá ser pior do que o melhor dos modelos de previsões. Isto significa que um modelo combinado nunca “perderá” para qualquer outro modelo, sempre prevendo melhor os dados.

Porém, isso só ocorre na teoria, porque infelizmente a eq. 2.24 não pode ser usada para encontrar o peso k , já que as variâncias e as covariâncias requeridas não são conhecidas. Podem ser estimadas a partir dos erros passados, mas geralmente existem poucos erros passados para realizar os cálculos. Outro empecilho é a inflexibilidade de k na combinação. Por exemplo, se f_i vier de um modelo simples de regressão e g_i de um modelo econométrico que seja continuamente atualizado e melhorado conforme aplicação das correções dos erros passados, seria razoável esperar que as previsões do modelo econométrico levassem pesos crescentes com o tempo conforme a melhora em suas previsões.

Para amenizar estes problemas, pode-se considerar um algoritmo onde o peso k_i se altera com o tempo, como equação abaixo:

$$2.26 \quad k_{i+1} = \frac{\sum_{t=n-m}^n (eg_t)^2}{\sum_{i=n-m}^n [(eg_t)^2 + (ef_t)^2]}$$

A eq. 2.26 é usada e as variâncias são estimadas a partir dos últimos m erros. Segundo Granger (1980), um valor apropriado para m seria 12 (quando se analisa dados mensais), embora esta seja uma escolha arbitrária. Se menos de doze erros estiverem disponíveis, a soma considerará apenas os erros conhecidos. Um peso apropriado para o caso em que não existem previsões anteriores é 1/2.

Essa técnica de combinação pode ser generalizada para mais de duas previsões, segundo a equação a seguir:

$$2.27 \quad c_i = k_{1,i}Z_{1,i} + k_{2,i}Z_{2,i} + k_{3,i}Z_{3,i} + \dots + k_{p,i}Z_{p,i}$$

Seja $e_{j,i}$ os erros do j -ésimo método de previsão. Então:

$$2.28 \quad A_{j,i} = \left[\sum_{i=n-m}^n (e_{j,i})^2 \right]^{-1}$$

$$2.29 \quad B_i = A_{1,i} + A_{2,i} + A_{3,i} + \dots + A_{p,i}$$

O peso do j-ésimo método deve ser:

$$2.30 \quad k_{j,i} = A_{j,i} / B_i$$

Entretanto, é melhor rejeitar todas as previsões realmente ineficientes, isto é, aquelas com valores particularmente pequenos de A e re-aplicar o procedimento anterior para combinar o restante.

Este método de combinação de previsões é denominado de *Combinação Ótima* que foca a minimização da variância dos erros de previsão combinada como o critério para a estimação dos pesos k, de cada previsão individual.

Para os leitores interessados em maiores detalhes na aplicação da combinação de métodos univariados na aplicação de carga elétrica, recomenda-se a leitura de Bou Issa (1996).