

Anderson da Silva Soares

**PREDIÇÃO DE SÉRIES TEMPORAIS ECONÔMICAS
POR MEIO DE REDES NEURAIS ARTIFICIAIS E
TRANSFORMADA WAVELET: COMBINANDO
MODELO TÉCNICO E FUNDAMENTALISTA**

Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica

Área de Concentração: Processamento de Sinais e Instrumentação
Orientadora: Profa. Dra. Maria Stela Veludo de Paiva

Dedicatória

Para Divino, Maristela, Gisele e Telma.

Agradecimentos

Primeiramente gostaria de agradecer a Prof. Dr. Maria Stela Veludo de Paiva, que me recebeu muito bem na USP, sempre dedicando muito carinho para com os seus alunos. Gostaria de agradecer também por depositar em mim a confiança de uma mudança de tema de pesquisa, sempre incentivando e apoiando com brilhantismo e motivação os trabalhos dos seus alunos.

Gostaria de expressar também minha profunda gratidão e reconhecimento a Universidade Católica de Goiás, que na pessoa do Professor Dr. Clarimar José Coelho investiram tempo e recursos em minha formação através da iniciação científica. Ainda guardo comigo toda a motivação, apoio e amizade construídas ao longo dos dois anos e meio de convivência com o Professor Dr. Clarimar, espero um dia poder recompensá-lo por todo o seu esforço e paciência.

Ao meus pais, Sr. Divino Benedito Soares e Sra. Maristela Correia da Silva Soares e minha irmã Gisele da Silva Soares, guerreiros e batalhadores. Cada esforço, cada boa nota, sempre tiveram por objetivo oferecer um futuro melhor a estes heróis.

A minha noiva Telma Woerle de Lima, amiga e companheira de todas as horas, sempre me apoiando nos momentos mais difíceis, inclusive financeiros. Momentos dos quais jamais esquecerei e espero torná-los imortais em breve.

Aos Professores Dr. Alexandre Cláudio Botazzo Delbem e Dr. Fernando Marques Federson do Instituto de Ciências Exatas e Computação da USP pelo incentivo e exemplo de amizade e solidariedade que a mim apresentaram. Os dias em São Carlos teriam sido mais tortuosos se não fossem os deliciosos churrascos na casa do Professor Dr. Alexandre.

Ao Professor Dr. Ivan Nunes da Silva pela excelência de qualidade nos cursos ministrados na pós-graduação, que em muito contribuíram para o desenvolvimento desta pesquisa.

Ao amigo e colega de curso Gustavo Teodoro Laureano, que desde a graduação é um grande amigo e companheiro científico de grande valia.

A Acta Brasil Informática que na pessoa do Sr. Danilo Alves Martins de Faria me concedeu um emprego no período em que estive sem bolsa de estudos.

À CAPES pela concessão da bolsa de mestrado e pelo apoio financeiro parcial no período de realização desta pesquisa.

A DEUS, pela vida.

“Eu guardei muitas coisas em minhas mãos, e perdi todas; mas todas que coloquei nas mãos de Deus, essas eu ainda possuo.”

– Martin Luther King, Jr.“

Resumo

Soares, A.S. Predição de Séries Temporais Econômicas por Meio de Redes Neurais Artificiais e Transformada Wavelet: Combinando Modelo Técnico e Fundamentalista. Dissertação de Mestrado - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 007.

Este trabalho apresenta um método de predição não linear de séries temporais econômicas. O método baseia-se na análise técnica e fundamentalista de cotação de ações, filtragem *wavelet*, seleção de padrões e redes neurais artificiais. No modelo técnico emprega-se a transformada *wavelet* para filtrar a série temporal econômica de comportamentos aleatórios ou não econômicos. Após a filtragem dos dados o algoritmo de projeções sucessivas é utilizado para a seleção de padrões de treinamento para a rede neural artificial, com o objetivo de selecionar os padrões de comportamento mais importantes na série. No modelo fundamentalista utiliza-se variáveis econômicas que podem estar correlacionadas com a série, com o objetivo de aprimorar a predição da série na rede neural artificial. Para avaliação do método são utilizados dados de séries temporais econômicas referentes à cotação de preços de ações negociadas na bolsa de valores de São Paulo, onde os resultados da predição do comportamento futuro são comparados com modelos matemáticos clássicos e com o modelo convencional, que se baseia somente na análise técnica. Apresenta-se uma comparação dos resultados entre modelos técnicos, modelos matemáticos e o método proposto. O modelo matemático utilizado (ARIMA) apresentou seu melhor desempenho em séries com pouca variância, porém com desempenho inferior quando comparado com o modelo técnico e com o método proposto. A avaliação do erro de predição em termos de RMSEP evidenciou que o método proposto apresenta os melhores resultados em relação aos demais métodos.

Palavras chave: Séries Temporais Econômicas, Transformada *Wavelet*, Redes Neurais Artificiais, Algoritmo de Projeções Sucessivas, Análise Técnica e Fundamentalista.

Abstract

Lista de Figuras

2.1	Rede neural recorrente NARMA.	8
3.1	Exemplo de funções base <i>wavelet</i>	18
3.	Transformada <i>wavelet</i> na forma de banco de filtros.	19
3.3	Decomposição <i>wavelet</i> de uma série temporal econômica.	0
4.1	Representação de um neurônio artificial.	4
4.	Principais funções de ativação utilizadas em neurônios artificiais.	4
4.3	Função sigmóide com diferentes valores de β (beta).	5
4.4	Rede neural artificial multi-camada não recorrente.	7
4.5	Rede Neural Artificial de Elman.	8
5.1	Visão geral da técnica proposta.	33
5.	Diagrama esquemático do modelo técnico.	34
5.3	Exemplo de seleção de variáveis utilizando o APS.	37
5.4	Diagrama esquemático do modelo fundamentalista.	38
6.1	Filtragem das séries utilizando transformada <i>wavelet</i>	44
6.	Resultado da seleção de padrões (Círculos vermelhos) por meio do algoritmo APS nas séries da Coteminas e Itaú.	45
6.3	Predição de séries econômicas com o método ARIMA e o método proposto.	48
C.1	Software de predição desenvolvido no MatLab.	69

C.	Exemplo de Resultado Obtido no Software de predição.	70
----	--	----

Lista de Tabelas

5.1	Variáveis fundamentalistas.	39
6.1	Resultado do coeficiente de Spearman para as variáveis fundamentalistas.	46
6.	Resultados RMSEP com técnicas clássicas e a técnica proposta. ¹ Método A: Modelo ARIMA, ² Método B: Modelo técnico, ³ Método C: Modelo técnico combinado com o modelo fundamentalista.	47

Lista de Abreviaturas e Siglas

ARMA - *AutoRegressive Moving Average*
ARIMA - *Autoregressive Integrated Moving Average*
BOVESPA - Bolsa de ValorES de São PAulo
DFT - *Discrete Fourier Transform*
IDFT - *Inverse Discrete Fourier Transform*
FGV - Fundação Getúlio Vargas
FIR - *Finite Impulse Response*
MLP - *Multi Layer Perceptron*
NARMA - *Non-linear Auto Regressive Mean Average*
PME - *Percentile Mean Error*
RPQME - *Root of Percentile Quadratic Mean Error*
RNA - Redes Neurais Artificiais
SRCC - *Spearman Rank Correlation Coefficient*
STFT - *Short Time Fourier Transform*
OLS - *Ordinary Least Squares*
TDNN - *Time Delay Neural Network*
VMF - Variáveis do Modelo Fundamentalista
VMT - Variáveis do Modelo Técnico
RPROP - *Resilient Propagation*

Lista de Símbolos

Escalares são representados em itálico; vetores e seqüências em minúsculo, matrizes em negrito.

- a : Ruído branco, um tipo de ruído produzido pela combinação simultânea de todas as frequências.
- b : fator de deslocamento de uma função wavelet.
- γ : Parâmetro de inclinação das funções sigmóide e hiperbólica.
- c : Fator de dilatação de uma função wavelet.
- η : Taxa de aprendizado da rede neural.
- E : Medida do erro total da rede.
- $g(\cdot)$: Função de ativação de um neurônio artificial.
- k : Intervalo de tempo entre duas observações.
- \cdot : Número de variáveis da matriz de treinamento X .
- p : Horizonte de predição da rede.
- ρ : Correlação de Spearman.
- ρ : Autocorrelação entre duas observações $x(t)$ e $x(t+k)$.
- \hat{A} : Operador polinômio autoregressivo.
- $\#$: Operador do polinômio de médias móveis.
- $^1_{rmsep}$: Média dos erros obtidos com um conjunto de testes.
- $^3_{rmsep}$: Desvio padrão dos erros obtidos com um conjunto de testes.
- T : Número de execuções de treinamento e predição de um conjunto de testes.
- x : Denota uma série temporal discreta.
- X : Denota uma série temporal contínua.
- \hat{x} : Valor predito pela rede neural artificial.
- X : Série temporal discreta representada no domínio da frequência.
- \bar{X} : Denota uma série temporal contínua representada no domínio da frequência.
- $\$$: Função de janelamento responsável pela delimitação do trecho da série.

\mathbb{R} : Denota o espaço dos números reais.

\mathbb{I} : Denota o espaço dos números imaginários.

H : Filtro passa alta.

L : Filtro passa baixa.

t : Unidade de tempo discreto.

1 : Média aritmética simples.

\mathcal{V}^2 : Variância da série temporal.

\downarrow : Operador de decimação.

μ : Limiar de ativação de um neurônio artificial.

u : Saída do combinador linear de um neurônio artificial.

y : Saída do neurônio artificial.

w : Peso sináptico de um neurônio artificial.

\hat{A} : Função escala.

\tilde{A} : Função wavelet-mãe.

Sumário

Resumo	vii
Abstract	ix
Lista de Figuras	xi
Lista de Tabelas	xiii
Lista de Abreviaturas e Siglas	xv
Lista de Símbolos	xvii
1 Introdução	1
1.1 Introdução ao Mercado de Ações	
1. Estrutura do Texto	4
2 Predição de Séries Temporais - Estado da Arte	5
.1 Considerações Iniciais	5
. Modelos Clássicos	5
.3 Técnicas de Predição Baseadas em Redes Neurais Artificiais	7
.4 Novas Tendências no Problema de Predição	9
.5 Considerações Finais	11
3 Transformada <i>Wavelet</i>	13
3.1 Considerações Iniciais	13
3. Perspectiva Histórica	13
3.3 Transformada de Fourier	14
3.4 Transformada <i>Wavelet</i>	16
3.4.1 Bases <i>Wavelets</i>	17
3.4. Análise Multiresolução	18
3.5 Considerações Finais	1

4	Redes Neurais Artificiais na Previsão de Séries Temporais Econômicas	23
4.1	Considerações Iniciais	3
4.	Fundamentos de Redes Neurais Artificiais	3
4.3	Arquiteturas de Redes Neurais Artificiais	6
4.3.1	Redes Neurais Artificiais Não Recorrentes	6
4.4	Redes Neurais Artificiais Recorrentes para Sistemas Variantes no Tempo	7
4.4.1	Rede Neural Artificial de Elman	8
4.5	Algoritmos de Treinamento de Redes Neurais Artificiais	9
4.5.1	Algoritmo <i>Resilient Propagation</i> (RPROP)	30
4.6	Considerações Finais	3
5	Metodologia	33
5.1	Considerações Iniciais	33
5.	Visão Geral da Técnica	33
5.3	Modelo Técnico	34
5.3.1	Algoritmo de Projeções Sucessivas	35
5.4	Modelo Fundamentalista	37
5.5	Predição: Combinando Modelo Técnico e Fundamentalista	40
5.5.1	Medidas de Avaliação	40
5.6	Software de Desenvolvimento	41
5.7	Considerações Finais	41
6	Resultados	43
6.1	Considerações Iniciais	43
6.	Suavização da Componente Aleatória das Séries	43
6.3	Seleção de Padrões Utilizando o APS	45
6.4	Combinando Modelo Técnico e Fundamentalista	46
6.5	Considerações Finais	49
7	Conclusões	51
7.1	Contribuições do Trabalho	51
7.	Publicações	5
7.3	Limitações	5
7.4	Trabalho Futuros	53
	Referências Bibliográficas	55
A	Layout do arquivo de Dados da BOVESPA	63

A.0.1	Cabeçalho	63
A.0.	Dados	64
B	Descrição das variáveis fundamentalista utilizadas	67
C	Software de Predição	69

Capítulo 1

Introdução

Este trabalho apresenta uma técnica de predição de séries temporais econômicas que faz uso de procedimentos econométricos, conhecidos como modelo fundamentalista, e de séries temporais, conhecidas como modelo técnico. O econométrico baseia-se em dados econômicos para a construção do modelo, incluindo no modelo de predição variáveis econômicas correlacionadas com a série estudada. O procedimento de séries temporais faz uso da transformada *wavelet* para o ajuste da série temporal e de RNAs para predição de valores futuros. Com a atenuação de movimentos não econômicos por meio da transformada *wavelet*, é possível estudar os efeitos econométricos baseando-se em dados econômicos para construir o modelo, empregando-se uma rede neural artificial como estimador de valores futuros.

As séries temporais são formadas por conjuntos de observações sequenciais ordenadas em relação ao tempo, sendo possível classificá-las quanto ao seu processo gerador dos dados, como estocásticas ou determinísticas (Hamilton, 1954; McCabe e Tremayne, 1995; Pereira, 1945).

As séries classificadas como determinísticas, possuem seu comportamento descrito por equações matemáticas conhecidas, sendo que a predição consiste na mera aplicação das equações que descrevem a série. Entretanto, a maioria dos problemas que envolvem séries temporais possui comportamento estocástico, e o modelo matemático que descreve o comportamento da série não existe ou não é conhecido. Como alternativa ao uso de modelos matemáticos, técnicas baseadas em reconhecimento de padrões como regressão e Redes Neurais Artificiais (RNAs) podem ser utilizadas no problema de predição de valores futuros de séries temporais de comportamento estocástico. Tais técnicas fazem uso da identificação de padrões históricos recorrentes, isto é, a identificação de padrões de oscilação na série temporal que aconteceram em instantes de tempo passados

$(t-1; t-2; \dots; t-n)$ para então prever os valores nos instantes futuros $(t+1; t+2; \dots; t+p)$ (Koskela et al., 1996; Kedem e Fokianos, 2000).

O início dos estudos no problema de predição de séries temporais estocásticas é atribuído a Weigend et al. (1990), que em 1966 propôs um modelo de predição que consistia de uma operação puramente linear onde a soma ponderada das observações prévias da série era utilizada para determinar valores futuros. A limitação de linearidade do modelo apresentado por Weigend foi superada pela de Harrison e Stevens (1976) que se baseia em um modelo probabilístico bayesiano (Spall, 1988; Pole et al., 1996). Entretanto, somente na metade da década de 80 é que os modelos probabilísticos como o de Harrison e Stevens foram amplamente difundidos devido à popularização dos computadores e consequentemente dos pacotes de softwares estatísticos utilizados neste tipo de modelo (Phillips, 1999; Pole, 1994).

Ainda na década de oitenta, foram iniciados estudos de técnicas de inteligência artificial, como as Redes Neurais Artificiais (RNAs) na predição de séries temporais. A habilidade das RNAs em extrair relações não lineares a partir de dados de entrada ruidosos, não lineares e não estacionários, tem produzido resultados melhores que os obtidos por procedimentos matemáticos e estatísticos (Kolen, 1994; O. Nerrand, 1993). Entretanto os avanços mais recentes no problema de predição de séries temporais econômicas fazem uso de conceitos não-relacionados a RNAs ou a modelos matemáticos e estatísticos. A partir dos estudos como o de Qi (1999), Racine (2001), Grôppo (2004) e Nagano et al. (2003) foi possível compreender que fatores econômicos são extremamente relevantes na compreensão e predição do comportamento de uma série temporal econômica, sendo que os padrões podem ser melhor compreendidos fazendo uso de dados econômicos.

1.1 Introdução ao Mercado de Ações

As séries temporais econômicas são um caso particular de séries temporais, distinguindo-se por conceitos econômicos relacionados ao comportamento da série e pela origem dos dados, que são publicados de forma oficial em períodos de tempo constantes (dia, mês, trimestre, etc.). Como exemplo é possível citar as séries econômicas de cotação de preços de ações, que são disponibilizadas de forma diária. As chamadas ações, que geram esses dados, representam a menor parcela do capital de uma empresa, podendo ser negociadas em mercados organizados como as Bolsas de Valores (Zancopé, 2000). As Bolsas de Valores então disponibiliza de forma

continua aos interessados, os dados referentes aos preços de negociação dessas ações.

Existem no mercado de capitais dois modelos de avaliação que procuram compreender e propor estratégias para a compra e venda de ações de uma determinada empresa: O modelo fundamentalista e o modelo técnico (Lameira, 2003).

O modelo fundamentalista analisa os aspectos econômicos e contábeis da empresa, bem como as variáveis macroeconômicas que tem influência direta ou indireta no valor da empresa. Como ilustração é possível citar a influência da cotação do dólar e do preço do barril de petróleo internacional no preço das ações de empresas petrolíferas e distribuidoras de combustível em geral, ou ainda, a influência da taxa de juros no valor das ações de bancos e instituições financeiras.

O modelo técnico baseia-se em dados históricos das ações da empresa no mercado, propondo modelos de predição a partir de padrões identificados na representação gráfica das variações que os preços demonstraram em um determinado período de tempo (Murphy, 1991; Tavares e Fonseca, 1987).

Os dois modelos de avaliação procuram, em essência, projetar o comportamento futuro dos preços das ações no mercado financeiro, sendo perfeitamente possível ainda a combinação das duas técnicas. Segundo Sá (1997), os dois modelos mencionados são complementares e a combinação deles poderá precisar de maneira mais adequada o comportamento dos preços.

Dow Jones pode ser considerado o primeiro estudioso das séries temporais econômicas de preços de ações. No início do século vinte, Dow propôs um modelo com o objetivo de descrever o comportamento de qualquer série através de três tipos de tendências (Brown et al., 1998): tendências de longo, médio e curto prazo. Segundo relata Brown et al. (1998), logo em seguida a proposta de Dow, Persons propôs um outro modelo baseado em quatro tipos de variações que descrevesse o comportamento de séries temporais econômicas. Os quatro tipos de variações propostos foram:

1. Tendência: variações caracterizadas por um movimento regular e contínuo de longo período de tempo, obedecendo uma direção geral.
2. Movimentos cíclicos: variações que correspondem aos chamados ciclos econômicos, ou seja, oscilações periódicas com desvios em torno da tendência.
3. Variações sazonais: as variações sazonais correspondem às oscilações de subida e de queda que ocorrem sempre em um mesmo período de tempo.

4. Variações Irregulares, residuais ou aleatórias: movimentos sistemáticos causados por fenômenos não econômicos de natureza aleatória, como por exemplo, catástrofes e boatos.

As quatro variações mencionadas seriam capazes de explicar qualquer série temporal econômica e, a partir destas considerações de Persons, esta composição da série temporal econômica tornou-se clássica, sendo considerada um marco no processo de estudos das séries. Alguns trabalhos recentes apresentam modelos matemáticos e estatísticos de ajustamento das variações sazonais a fim de explicar a série somente a partir de parâmetros econômicos (Al-Deehani, 2006; Al-Saad, 2005). Em países desenvolvidos, a suavização de movimentos sazonais e irregulares em séries temporais econômicas já é considerado uma prática oficial. Entretanto, segundo Hotta (1988), a utilização deste tipo de técnica em países em desenvolvimento como o Brasil, deve ser feita com maior cuidado, em razão de estarem submetidos frequentemente a fortes mudanças estruturais e econômicas, gerando grandes variações no comportamento das séries. As técnicas de suavização tradicionais fazem uso de ferramentas como média móvel, alisamento exponencial e modelos ARIMA (*Autoregressive Integrated Moving Average*), sendo que o modelo ARIMA será descrito no Capítulo 2. Essas técnicas de ajustamento são importantes para que seja possível a utilização de um estimador eficiente da série temporal, uma vez que a suavização funciona como um filtro da série.

1.2 Estrutura do Texto

No Capítulo 1 apresenta-se uma revisão bibliográfica dos principais trabalhos de predição de séries temporais encontrados na literatura. No Capítulo 2 é feita uma revisão sobre a transformada *wavelet* e seu funcionamento como filtro em séries temporais. No Capítulo 3 é apresentada uma revisão sobre redes neurais artificiais e sua aplicação para tratar sistemas variantes no tempo como as séries temporais. No Capítulo 4 é descrito o método proposto. No Capítulo 5 são apresentados os resultados obtidos e finalmente no Capítulo 6, Ao término do Capítulo, são tecidas considerações sobre as contribuições, limitações e continuidades do trabalho para o problema apresentado.

Capítulo 2

Predição de Séries Temporais - Estado da Arte

2.1 Considerações Iniciais

Este Capítulo apresenta os principais trabalhos sobre predição de séries temporais. Técnicas clássicas como ARIMA e redes neurais artificiais são descritas, em um contexto de evolução histórica do tratamento do problema de predição de séries temporais.

2.2 Modelos Clássicos

Em 1970, os professores George E. P. Box e Gwilym M. Jenkins publicaram o livro, *Time Series Analysis: forecasting and control* (Box et al., 1976). O modelo de Box & Jenkins, também conhecido como auto-regressivo integrado de média móvel - ARIMA (*AutoRegressive Integrated Moving Average*), foi elaborado a partir da coletânea de diversos trabalhos publicados na época, trazendo como mérito, a unificação das principais técnicas estatísticas, em uma única metodologia para construir modelos de predição de séries temporais.

O modelo de Box & Jenkins assume que qualquer série temporal é um processo estocástico estacionário ¹ gerado pela passagem sucessiva de um processo ruído branco a_t ² por modelos de

¹Uma série temporal é dita estacionária quando não existe tendência de crescimento ou decrescimento dos dados, ou seja, os dados oscilam sobre uma média constante com variância constante.

²Um processo estocástico é chamado de ruído branco quando o valor tomado no tempo t , independe do valor tomado no tempo $t - k$, não apresentando qualquer dependência temporal.

médias móveis (equação .1) e auto-regressivos (equação .).

$$x_t = \mu + a_t - \theta_{(1)}a_{(t-1)} - \theta_{(2)}a_{(t-2)} \dots - \theta_{(q)}a_{(t-q)} \quad (.1)$$

$$x_t = \phi_1 x_{(t-1)} + \phi_2 x_{(t-2)} + \dots + \phi_p x_{(t-p)} + a_t \quad (.)$$

onde ϕ são os parâmetros da estrutura auto-regressiva de ordem p , θ são os parâmetros da estrutura médias móveis de ordem q , μ é a média da série e a_t representa o ruído branco com média zero e variância σ^2 . De acordo com Box e Jenkins (1976), qualquer processo estocástico pode ser representado por um modelo auto-regressivo e/ou médias móveis. No caso de processos estocásticos não estacionários, aplica-se o operador de diferenciação para tornar a série estacionária.

O artigo de Harvey e Todd (1983) foi um dos principais artigos da década de oitenta voltado especificamente para o problema de predição de séries temporais econômicas. Entretanto, seu trabalho era limitado na generalização do modelo ARIMA proposto por Box-Jenkins para o problema de predição de séries temporais econômicas (Granger e Morris, 1976). No artigo, Harvey e Todd (1983) demonstraram que ao separar as quatro componentes de séries temporais econômicas propostas por Down Jones, o modelo ARIMA poderia ser aplicado ao problema de predição das componentes de tendência e sazonalidade da série. Para separar as componentes, utilizaram o filtro de Kalman para suavizar o comportamento aleatório, onde a componente de tendência e sazonalidade eram de interesse. Os resultados mostraram que a predição ocorria com certo sucesso para a componente de sazonalidade, mas não apresentava resultados satisfatórios para a componente de tendência.

Entretanto, os trabalhos de McCabe e Tremayne (1995) e Leybourne et al. (1996) questionam a utilização do filtro de diferenciação no modelo ARIMA para séries não estacionárias. Utilizando séries temporais econômicas e testes estatísticos de hipótese, os resultados dos dois trabalhos mostraram que para várias séries temporais do tipo estocásticas, a série gerada após a aplicação de diferenciação, pode não ser da mesma fonte, ou seja, gera-se uma série estatisticamente diferente da série original e conseqüentemente seu modelo de predição pode ser rejeitado.

2.3 Técnicas de Predição Baseadas em Redes Neurais Artificiais

A primeira aplicação que utilizou RNA como estimador, ao invés de modelo matemático como o ARIMA, foi a de White (1988). Apesar da ausência de comparação com os modelos clássicos como o ARIMA, o artigo foi importante na discussão de técnicas inteligentes na detecção dos padrões presentes nas séries temporais econômicas. Como dados de testes foi utilizado somente a série econômica dos preços de cotações da empresa IBM, e o algoritmo OLS (*Ordinary Least Squares*) para treinamento da RNA. O autor conclui que a aplicação de técnicas inteligentes como as RNAs ainda careciam principalmente de algoritmos de treinamentos eficientes para o aprendizado de regularidades não lineares presentes nas séries temporais econômicas.

No ano de 1991 a comunidade científica conheceu dois trabalhos notoriamente reconhecidos até os dias atuais, como marcantes no problema de predição de séries temporais econômicas. A importância desses resultados deve-se à consistência dos resultados obtidos pelos trabalhos de Jang (1991) e Matsuba (1991). A aplicação de Jang (1991) conseguiu prever com erros pequenos algumas séries econômicas utilizando duas RNAs do tipo MLP (*Multi Layer Perceptron*) com algoritmo de treinamento *backpropagation*. Uma rede foi treinada para prever a tendência de preços a longo prazo e a segunda rede foi treinada para aprender a tendência a curto prazo. Os dados usados foram valores de pico, valores de baixa, fechamento e o volume negociado para cada dia. Os autores concluíram que a rede neural implementada apresentou uma resposta mais satisfatória do que técnicas estatísticas baseadas em regressão. Já o trabalho de Matsuba empregou uma rede neural do tipo *feedforward* com os n últimos valores de índice como entrada na rede, e os p preços futuros como saída da rede, empregando também duas redes neurais, uma aprender as características globais e outra para aprender as características locais ou pequenas flutuações para realizar a predição de séries temporais.

Outro importante trabalho foi o de Connor et al. (1994), que propôs uma técnica mais sofisticada para o problema de predição de séries temporais. A proposta consistiu na filtragem do comportamento aleatório por meio da identificação e filtragem de *outliers*³, e da estimação da série a partir dos dados filtrados por meio de uma RNA MLP recorrente. A principal característica da RNA proposta por Connor, é possuir laços de realimentação dos erros produzidos pela saída da rede conforme ilustra a Figura .1.

A Figura .1 mostra uma RNA recorrente para aproximação de um modelo ARMA não linear

³Uma observação é classificada como *outlier* quando não segue o mesmo padrão de dependência espacial como a maioria das observações do conjunto de dados.

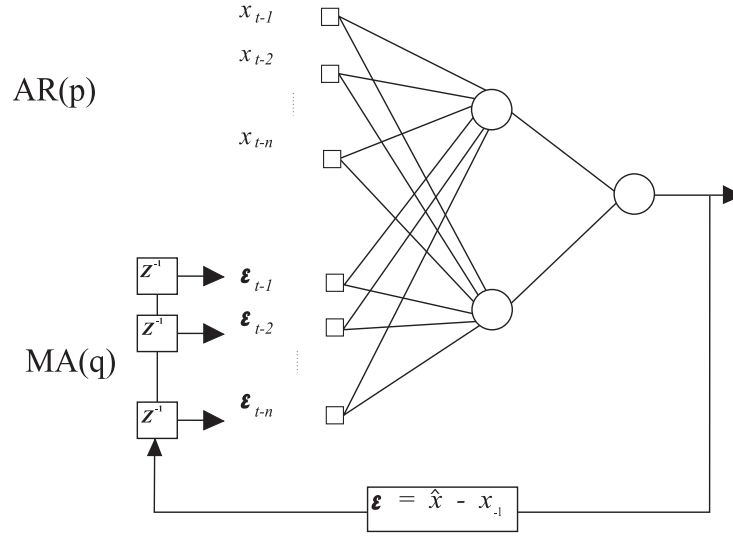


Figura .1: Rede neural recorrente NARMA.

conhecido como NARMA (*Non-linear Autoregressive Moving Average*), que faz uso do modelo auto-regressivo (AR) de ordem p e do modelo média-móvel de ordem q , onde a saída \hat{x} é dada pela equação .3.

$$\hat{x}(t) = \sum_{i=1}^m W_i g \left\{ \sum_{j=1}^p w_{ij} x_{(t-j)} + \sum_{j=1}^q \bar{w}_{ij} (x_{(t-j)} - \hat{x}_{(t-j)} + \mu_i) \right\} \quad (.3)$$

onde μ é o limiar de ativação do i -ésimo neurônio, g é a função de ativação, x é a entrada e \hat{x} a saída da rede, W_i são os pesos entre os neurônios da camada intermediária (ou escondida) e de saída da rede, w_{ij} são os pesos entre as entradas externas e a camada escondida, \bar{w}_{ij} são os pesos entre a primeira camada de neurônios com entradas de retro-alimentação (recorrentes) e a camada de neurônios escondidos, μ_i é o bias que modifica as saídas dos neurônios. Foi utilizada a série temporal de demanda de consumo de energia elétrica, onde o problema consistia na predição do consumo no tempo futuro. Os autores concluíram que a proposta apresenta melhores resultados do que as redes convencionais na presença de *outliers*.

A proposta de Connor foi a primeira de inúmeros trabalhos que propuseram a combinação de RNAs com modelos clássicos. Outros trabalhos similares foram as propostas de Wang e Leu (1996), Diniz et al. (1999) e Temizel e Ahmad (2005). Os três artigos propuseram um sistema preditivo que faz uso de RNAs e do modelo ARIMA. O algoritmo consistiu no uso do modelo ARIMA na etapa de pré-processamento da série, onde aplicou-se o modelo ARIMA por meio

de operações de diferenciação e posteriormente a predição da série. O trabalho de Wang e Leu (1996) foi aplicado na bolsa de valores de Taiwan enquanto que o de Diniz na bolsa de valores de São Paulo - Brasil. Os resultados dos dois trabalhos mostraram que a RNA melhorou sua capacidade preditiva das séries após a combinação com o modelo ARIMA, pois, a diferenciação da série elimina alguns tipos de comportamentos não estacionários.

Apesar do relativo sucesso da utilização de técnicas matemáticas e estatísticas na fase de pré-processamento, nos trabalhos de Aussem e Murtagh (1997) e Malliavin e Mancino (2000), a utilização de técnicas estatísticas, combinadas ou não com RNAs, começou a ser confrontada com técnicas como transformadas de Fourier e wavelets, baseadas no domínio da frequência. A comparação dos métodos mostrou que as técnicas no domínio da frequência tiveram resultados superiores quando comparadas com as técnicas estatísticas, haja vista que no domínio transformada é possível evidenciar características e comportamentos não notados no domínio original dos dados. Desde então, a maioria dos trabalhos nos últimos dez anos vem propondo a utilização de métodos no domínio da frequência.

Saad et al. (1998) apresentou um estudo comparativo de três tipos de arquiteturas de RNAs para o problema de predição de séries temporais econômicas. Foram estudadas as arquiteturas com atraso de tempo (TDNN - *Time Delay Neural Network*), recorrente, e probabilística. As RNAs foram testadas na predição de tendências de curto prazo da bolsa de valores de Nova York, onde os resultados demonstraram que as três arquiteturas estudadas têm uma capacidade preditiva semelhante, com vantagem para a arquitetura TDNN, por sua facilidade de implementação computacional. Entretanto, o trabalho de Koskela et al. (1996) apresenta o estudo comparativo entre RNAs do tipo MLP (*Multi Layer Perceptron*), FIR (*Finite impulse response*) e rede de Elman, mostrando que utilizando o algoritmo *Backpropagation* para as três arquiteturas de redes, a rede de Elman apresentou uma memória de longo prazo melhor, e conseqüentemente uma predição melhor no problema de séries temporais.

2.4 Novas Tendências no Problema de Predição

Os trabalhos de Dhatt et al. (1999a), Gurgul et al. (2003), Grôppo (2004) e Cajueiro e Tabak (2005), apesar de não tratarem especificamente de predição, trouxeram um novo foco na abordagem do problema de predição de séries temporais do tipo econômica. Dhatt estudou a relação entre as variáveis fundamentalistas e o comportamento de preços das ações a partir da

bolsa de valores da Korea. Dhatt agrupou um grupo de ações de um mesmo setor de mercado e propôs a teoria de que todas as ações do grupo setorial eram influenciadas pelas variáveis econômicas daquele setor. Para isso, provou de forma experimental, por meio de um modelo de regressão linear múltipla, que as variáveis econômicas específicas do setor econômico que as ações estavam inseridas, influenciavam no comportamento dos seus preços.

Gurgul também estudou os efeitos de variáveis fundamentalistas no comportamento da série temporal de preços de ações. Entretanto, o estudo foi limitado à variável de taxa de lucro da empresa, empregando um modelo univariado, ao contrário de Dhatt. Utilizando o teste estatístico de *t-student*, verificava-se se a taxa de lucro da empresa influenciava no comportamento da série econômica. A conclusão dos autores sugere que os preços geralmente absorvem a divulgação da taxa de lucro da empresa de forma rápida, e dentro do mesmo dia a notícia influencia fortemente a formação da série, e perpetua por vários dias ou até meses.

Em âmbito nacional, dois trabalhos recentes, com dados do Brasil, mostram também a tendência da nova abordagem. Cajueiro e Tabak (005) fizeram o estudo das possíveis fontes do comportamento de longo alcance no mercado financeiro brasileiro. Utilizando o expoente de Hurst e o coeficiente de Spearman, o autor afirma que encontrou a evidência de que variáveis específicas da empresa emissora dos títulos de ações na bolsa podem explicar, pelo menos parcialmente, comportamento de longo prazo, representando um importante avanço nos problemas relacionados a séries temporais econômicas. Entretanto, na conclusão, faz a advertência que os preços não estão dirigidos unicamente por variáveis fundamentalistas, mas também por outras características do mercado, como o comportamento especulativo (no contexto deste trabalho, o comportamento especulativo refere-se a análise técnica) e as chamadas “bolhas especulativas” têm papéis significativos na determinação dos preços, principalmente em mercados emergentes como o Brasil.

Grôppo (004) estudou a relação entre um conjunto de variáveis macroeconômicas e o mercado acionário brasileiro, utilizando um enfoque multivariado com o método de auto regressão vetorial com correção de erro. O autor buscou analisar o efeito das variáveis macroeconômicas sobre o índice da Bolsa de Valores de São Paulo, e conseqüentemente de suas séries econômicas. O período analisado compreendeu os meses de janeiro de 1995 a dezembro de 003. O autor concluiu que existem relações significativas entre as variáveis econômicas, taxa de câmbio e a taxa de juros de curto prazo, com o comportamento das séries que compõe o índice Bovespa. Os resultados obtidos deixam claro, segundo o autor, a elevada sensibilidade das séries econômicas frente à taxa de juros

real de curto prazo (SELIC). Por sua vez, a cotação do petróleo no mercado internacional não explica significativamente o comportamento das séries. O autor também concluiu que o próprio índice Bovespa também tem um grande poder explicativo sobre as séries temporais econômicas que compõe o índice.

E finalmente, Lam (004) foi a primeira a propor uma RNA que integrasse os modelos técnicos e fundamentalistas. Fazendo uso de vinte e sete variáveis fundamentalistas, Lam demonstrou que uma RNA do tipo MLP com algoritmo de treinamento *backpropagation* é capaz de integrar os dois modelos. Foram utilizados dados de empresas com ações cotadas na bolsa de valores de Nova York, no período de 1985 a 1995. Para a integração dos modelos, foi utilizado a técnica estatística de planejamento de experimentos, detalhada em Mead (1990). Com os resultados, a autora concluiu que a habilidade de predição da rede é melhorada para a maioria das séries estudadas.

2.5 Considerações Finais

Este Capítulo apresentou a revisão bibliográfica das principais técnicas encontradas na literatura. Foi descrita uma evolução histórica dos métodos de tratamento do problema de predição de séries temporais em geral, até os trabalhos mais recentes. Estes artigos recentes evidenciam a importância de utilizar as abordagens fundamentalista e técnica em conjunto no problema de predição de séries temporais econômicas. Contudo, a combinação desses modelos ainda é tímida quando tratada no escopo de técnicas de inteligência artificial como as RNAs.

Capítulo 3

Transformada *Wavelet*

3.1 Considerações Iniciais

Neste Capítulo é feita uma introdução a transformada *wavelet* e seu funcionamento como filtro no problema de predição de séries temporais.

3.2 Perspectiva Histórica

A transformada *wavelets*, ou análise *wavelet*, mostra muitas origens diferentes, como ilustra Meyer (1993a). Entretanto, sua idealização foi concebida no trabalho de Haar, que no apêndice de sua tese de doutorado, demonstrou que um sinal qualquer poderia ser diferenciado e integrado de modo a conservar sua energia inicial. Por volta de 1930, vários grupos trabalhando independentemente publicaram trabalhos desenvolvidos com a teoria *wavelet* proposta por Haar. É possível citar, por exemplo, o trabalho de Paul Levy descrito em Meyer (1993b), que investigou o movimento Browniano usando a base de *wavelets* de Haar, mostrando que as funções da base de Haar eram melhores do que as da base de Fourier para estudar os detalhes do movimento Browniano.

Contudo, somente em meados da década de oitenta que seu uso foi amplamente aceito e adotado pela comunidade científica. Parte dessa aceitação deve-se ao ganho de poder computacional dos computadores e aos trabalhos de Daubechies (1990), Mallat (1989b) e Morlet et al. (1989).

A partir da década de noventa é possível observar uma tendência da comunidade científica em explorar o potencial das *wavelets* em aplicações de codificação, representação, compressão

e filtragem de sinais. Paralelamente aos trabalhos voltadas para aplicações específicas, diversas contribuições de novos algoritmos e estratégias de implementações computacionais propuseram versões para a transformada rápida discreta (utilizando-se bancos de filtros) e algoritmos melhorados para resolver efeitos de bordas e particionamento do espectro do sinal, de forma a aumentar a resolução nas faixas de frequências mais altas, utilizando-se técnicas como os "pacotes de *wavelets*" (*wavelet-packets*), e multiwavelets.

O início dos estudos de sinais no domínio da frequência é atribuída a Joseph Fourier, que em 1807 iniciou o estudo das teorias de análise de frequência a partir das séries de Fourier, onde qualquer sinal poderia ser representado pela soma infinita de senos e co-senos. Explorando as séries de Fourier, os matemáticos foram gradualmente se direcionando a noção prévia de análise de frequência e escala conduzindo a famosa transformada de Fourier para análise de sinais no domínio da frequência.

3.3 Transformada de Fourier

A transformada de Fourier é tradicionalmente utilizada para se fazer a transformação tempo-frequência (Gencay et al., 2001; Soltani, 2000). A transformação dos dados para o domínio da frequência mostrar a mesma informação em outro domínio, tornando algumas características mais evidentes no domínio transformado, como por exemplo a componente aleatória da série temporal (Ingle e Proakis, 2000).

Dado um conjunto de pontos x no domínio do tempo, a transformada de Fourier consiste em determinar o conjunto de pontos X correspondentes no domínio da frequência. Isso permite identificar os aspectos mais frequentes e comuns do sinal. Permite ainda, filtrar, eliminar ou ampliar com maior facilidade uma faixa específica de frequência (Attia, 1999; Bracewell, 1988; Claerbout, 1999; Clarke, 2000).

A transformada de Fourier é calculada pela Equação (3.1).

$$\bar{X}(f) = \int_{-\infty}^{+\infty} x(t)e^{-j2\pi ft} dt; \quad (3.1)$$

onde $j = \sqrt{-1}$ e f é a frequência do sinal no domínio do tempo. A transformada inversa de Fourier que recupera a função original x a partir de \bar{X} é definida como:

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df; \quad (3.)$$

A partir do conjunto de pontos finitos obtidos no processo de discretização é possível aplicar a fórmula de aproximação dos coeficientes da transformada de Fourier conhecida como transformada discreta de Fourier (*Discrete Fourier Transform*, DFT) (Elliott e Rao, 198).

A DFT de N pontos é definida como:

$$X(f) = \frac{1}{N} \sum_{n=0}^{N-1} x(t) e^{-\frac{jn2\pi f}{N}} \quad (3.3)$$

A transformada discreta de Fourier inversa (*Inverse Discrete Fourier Transform*, IDFT) recupera o sinal do domínio da frequência para o domínio do tempo. A IDFT é definida como:

$$x(t) = \sum_{n=0}^{N-1} X(f) e^{\frac{jn2\pi f}{N}}; \quad (3.4)$$

A divisão da série temporal em segmentos de modo a permitir localização temporal de frequência é uma forma de melhorar o desempenho da transformada de Fourier. A localização temporal de frequências é obtida através do janelamento da série x em instantes de tempo t distintos. A transformada de Fourier é aplicada a segmentos definidos por uma janela definida como $\$$. O procedimento é conhecido como transformada de Fourier janelada (*Short Time Fourier Transform*, STFT) (Kaiser, 1994). A STFT é definida por:

$$F(p; f) = \sum_{t=0}^{N-1} x_t \$_{(t-p)} e^{-ift} \quad (3.5)$$

onde $\$$ é a função de janelamento responsável pela delimitação do trecho considerado do sinal. A posição da janela dentro do sinal é dada por p .

A dificuldade para definição do comprimento da janela e o fato da janela permanecer constante ao longo do sinal constituem a principal desvantagem para o emprego da STFT, desvantagens que sugerem o emprego da transformada *wavelet*.

3.4 Transformada *Wavelet*

A transformada *wavelet* utiliza janelas com comprimentos variáveis que se ajustam melhor às características do sinal em contraste com a STFT. A grande vantagem da transformada *wavelet* é a possibilidade da análise tempo-frequência que permite a localização de altas frequências em intervalos de tempo específicos, típico de séries temporais econômicas, tornando possível a suavização do comportamento aleatório da série em trechos específicos. A transformada *wavelet* para a análise e filtragem de séries temporais tem sido proposta em diversos trabalhos nos últimos anos (Renaud et al., 005; Fryzlewicz, 003) em substituição às técnicas clássicas como a transformada de Fourier.

Em sua versão discreta a transformada *wavelet* é dada por:

$$f(t) = \sum_{k=-\infty}^{\infty} d_k A_k(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} c_k^j \tilde{A}_k^j(t) \quad (3.6)$$

onde a função

$$\tilde{A}_k^j(t) = \tilde{A}(2^j t - k) \quad (3.7)$$

é chamada de *wavelet* mãe e a função

$$\tilde{A}_{j,k}(t) = 2^{\frac{j}{2}} \tilde{A}(2^j t - k) \quad (3.8)$$

é chamada de *wavelet* pai ou função escala e os coeficientes j e k são os coeficientes de transformação *wavelet*.

A transformada *wavelet* especifica a localização (via translação) e frequência (via dilatação ou compressão) utilizando a *wavelet* pai e mãe respectivamente.

$$\tilde{A}_{c,b}(t) = \frac{1}{\sqrt{c}} \tilde{A}\left(\frac{t-b}{c}\right) \quad (3.9)$$

$$\tilde{A}_{c,b}(t) = \frac{1}{\sqrt{c}} \tilde{A}\left(\frac{t-b}{c}\right) \quad (3.10)$$

Onde os parâmetros c e b variam continuamente no espaço em \mathbb{R} , de modo a obter a *wavelet*

dilatada ou comprimida pelo fator c e deslocada pelo fator b . Através de compressões/dilatações e deslocamentos das funções *wavelets* é possível aproximar qualquer sinal nos planos \mathbb{R} e \mathbb{I} .

Para $c > 1$ têm-se o efeito de dilatação da função *wavelet* e para $c < 1$ têm-se o efeito de contração da função *wavelet* sendo possível analisar os aspectos globais e locais das séries. À medida que b varia, a função f é analisada localmente em torno de b .

3.4.1 Bases *Wavelets*

Enquanto a transformada de Fourier utiliza funções trigonométricas seno e cosseno, a transformada *wavelet* pode utilizar qualquer função matemática que satisfaça os critérios de admissibilidade expressos nas equações 3.11 e 3.1 .

$$\int_{-\infty}^{\infty} \tilde{A}(t) dt = 0 \quad (3.11)$$

$$\int_{-\infty}^{\infty} |\tilde{A}(t)|^2 dt = 1 \quad (3.1)$$

A equação 3.11 expressa que para uma função ser *wavelet*, sua integral deve ser zero. Isso garante que a função *wavelet* tenha uma forma do tipo onda, e a equação 3.1 define que a função deve ter energia unitária, garantindo que a função tenha decaimento rápido para a localização temporal (Hernandez e Weiss, 1996; Mallat, 1999).

A *wavelet* de Haar foi a primeira função conhecida com comportamento *wavelet* e também a mais simples em termos matemáticos. A *wavelet* de Haar é dada pela equação 3.13.

$$\tilde{A}(t) = \begin{cases} 1; & t \leq \frac{1}{2} \\ -1; & \frac{1}{2} \leq t < 1 \\ 0; & t > 1 \end{cases} \quad (3.13)$$

A base de Haar tem o inconveniente de ser formada por funções descontínuas, expressas na equação 3.13, entretanto sua facilidade de compreensão e implementação motivam o uso da base de Haar como *wavelet* mãe.

A partir da base de Haar outras bases foram desenvolvidas na forma de família de funções similares, como por exemplo a família de bases de Daubechies (Daubechies, 1990), Meyer (Meyer,

1993a) e Coiflet (Huang e Hsieh, 00). A Figura 3.1 ilustra a representação dessas bases.

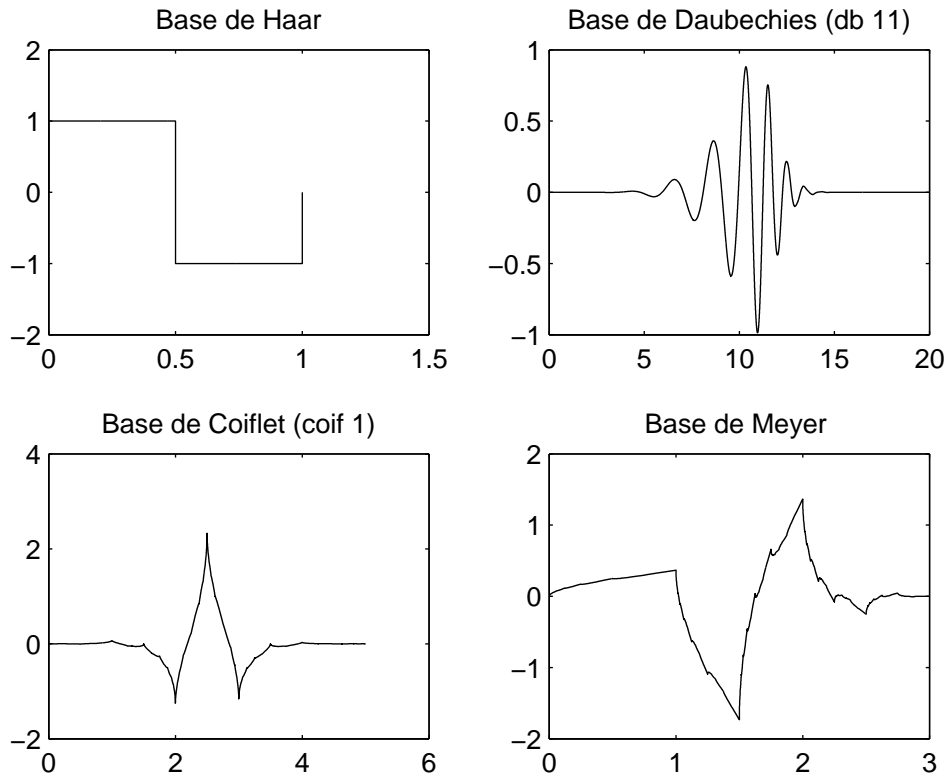


Figura 3.1: Exemplo de funções base *wavelet*.

Uma questão que sempre surge na aplicação da análise *wavelet* é a escolha da base *wavelet* mais adequada para analisar um certo tipo de sinal. Ainda não existe uma receita para esse procedimento, apenas recomendações úteis, como por exemplo, a forma da base *wavelet* escolhida deve refletir as características do sinal. A escolha da base *wavelet* é um fator importante, mas a escolha de uma base adequada ainda é tema de pesquisa em aberto. Outra alternativa é definir experimentalmente a base *wavelet* a ser utilizada no problema em estudo.

3.4.2 Análise Multiresolução

A representação de um sinal em diferentes níveis de detalhamento é conhecida por multiresolução. Os níveis de detalhamento são obtidos empregando-se a transformada *wavelet* através de filtros, técnica desenvolvida por Mallat (Mallat, 1989a). A aproximação de uma função x no domínio t é efetuada obtendo sua aproximação grosseira mais os detalhes em diferentes níveis, ou seja, com aproximações e níveis de detalhes aplicados de forma recursiva, permitindo analisar sinais em múltiplas bandas de frequências (Mertins, 1996).

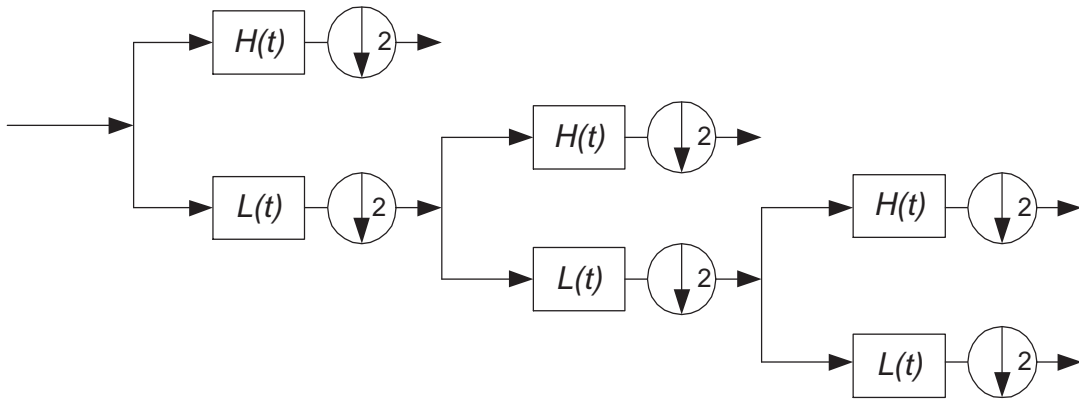


Figura 3. : Transformada *wavelet* na forma de banco de filtros.

A transformada *wavelet* em forma de banco de filtros ilustrada na Figura 3. , utiliza um filtro passa-baixa L e outro passa-alta H , onde o filtro passa-alta produz os coeficientes de detalhes da decomposição *wavelet* o filtro passa-baixa produz os coeficientes de aproximação da decomposição *wavelet*. A transformada funciona como um filtro de banda passante para cada nível de decomposição (Mertins, 1996). O sinal é decomposto simultaneamente utilizando os filtros H e L , repetidindo-se a decomposição no canal L de modo a incrementar a resolução tempo-freqüência até um nível n .

A decomposição *wavelet* tem como característica filtrar e decimar o sinal. Na Figura 3. o operador \downarrow de decimação diminui a taxa de amostragem do sinal, ou seja, elimina pontos intercalados. Decimar um sinal pelo fator 2 significa que o sinal será sub-amostrado por um fator de 2, obtendo-se metade do sinal original pela redução da taxa de amostragem. A Figura 3.3 apresenta um exemplo da aplicação da decomposição *wavelet* ilustrada na Figura 3. .

Na Figura 3.3(a) é mostrada uma série temporal obtida das cotações de preço das ações do grupo Pão de Açúcar no período de Janeiro a Novembro de 2005. A série foi submetida à decomposição *wavelet* utilizando-se a base de Daubechies (db 11), obtendo-se a aproximação a_1 e seus detalhes d_1 para um primeiro nível de resolução, aproximação a_2 e detalhes d_2 para o segundo nível de resolução, sendo possível analisar a série em dois níveis de decomposição.

Para muitos sinais, inclusive as séries temporais econômicas, as componentes de baixa freqüência são aquelas que contêm maior informação. Já nas componentes de alta freqüência, são encontrados detalhes que contêm a maior parte do comportamento aleatório da série, possibilitando o emprego de um limiar de corte na componente de detalhes.

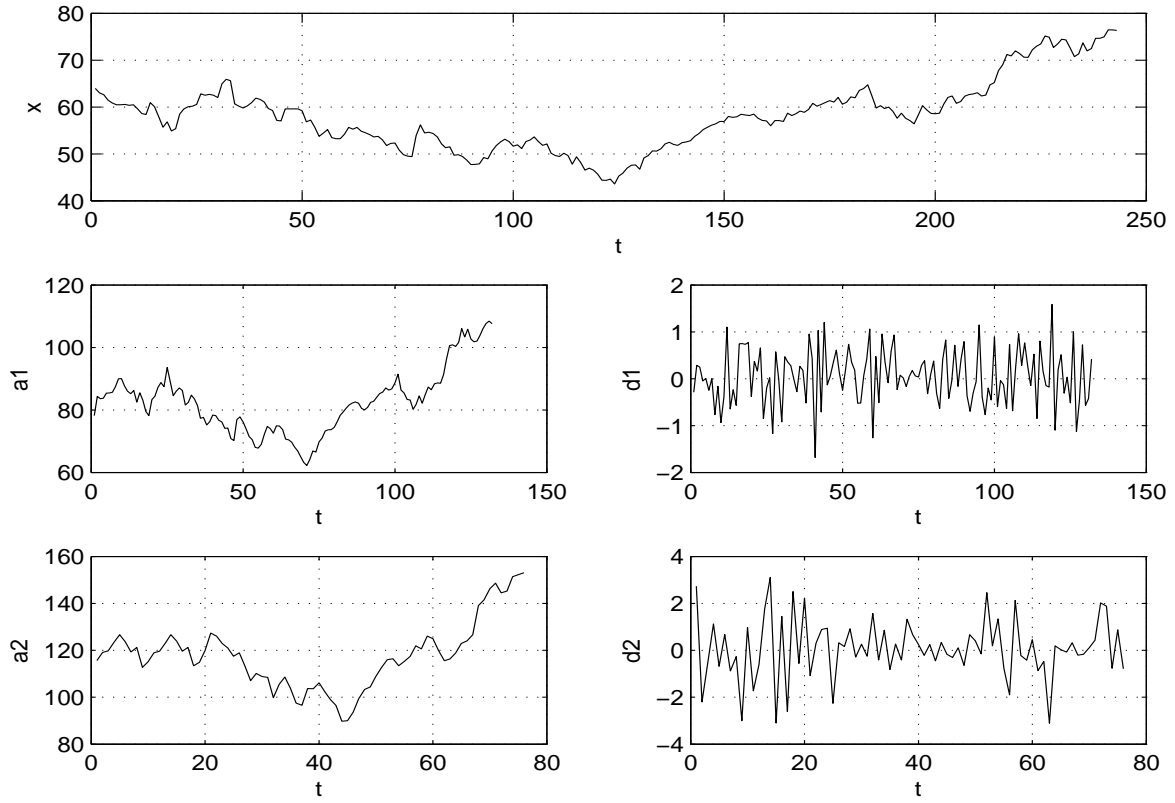


Figura 3.3: Decomposição *wavelet* de uma série temporal econômica.

A equação 3.14 apresenta o limiar universal proposto por Donoho e Downie (Donoho, 1995; Downie e Silverman, 1998), para o corte dos coeficientes de detalhe para cada nível de resolução na transformada *wavelet*.

$$\frac{3}{4}\sqrt{2\log(n)} \quad (3.14)$$

sendo $\frac{3}{4}$ o desvio padrão e n o comprimento dos dados.

A equação 3.14 é aplicada como limiar de corte nos coeficientes de detalhes para cada nível de resolução na transformada *wavelet*. Utilizando a transformada *wavelet* na série temporal é possível realçar ou suavizar determinadas características em trechos localizados, o que não era possível utilizando técnicas como transformada de Fourier ou médias móveis.

3.5 Considerações Finais

Neste Capítulo foram apresentadas as principais técnicas de processamento de séries temporais no domínio da frequência. Como mencionado na seção 3.4, a transformada *wavelet* para a análise e filtragem de séries temporais tem sido proposta nos últimos anos em substituição às técnicas clássicas como a transformada de Fourier, cujas limitações foram apresentadas na seção 3.3.

Capítulo 4

Redes Neurais Artificiais na Previsão de Séries Temporais Econômicas

4.1 Considerações Iniciais

Neste Capítulo são apresentados alguns conceitos fundamentais sobre as redes neurais artificiais, princípio de funcionamento e a motivação no uso das RNAs no problema de previsão de séries temporais econômicas. Apresenta-se algumas das principais arquiteturas de RNAs utilizadas no problema de predição de séries temporais, concluindo-se com uma comparação sobre arquiteturas multi-camada recorrente e multi-camada não recorrente.

4.2 Fundamentos de Redes Neurais Artificiais

RNAs são modelos computacionais não lineares, que de forma similar ao cérebro humano, são organizadas através de elementos básicos de processamento denominados neurônios artificiais, que se interligam formando redes capazes de processar, armazenar e transmitir informações vindas de um meio exterior (Haykin, 2001). O modelo geral de um típico neurônio artificial conhecido como neurônio de Mcculloch-Pitts (Haykin, 2001) é ilustrado na Figura 4.1, onde o neurônio artificial, de forma similar a um neurônio biológico, recebe informações de outros neurônios ou de um meio exterior e as processa de modo a oferecer uma resposta em sua saída.

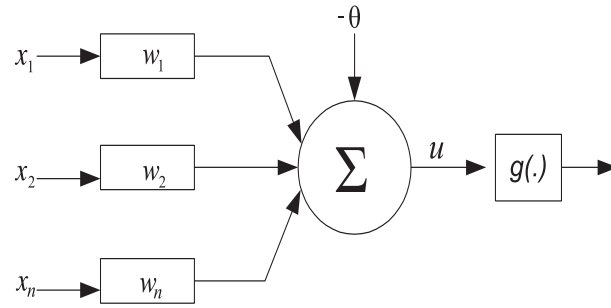
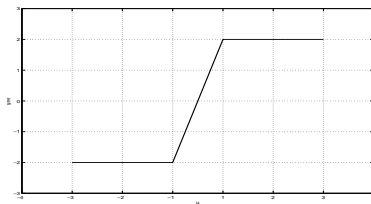
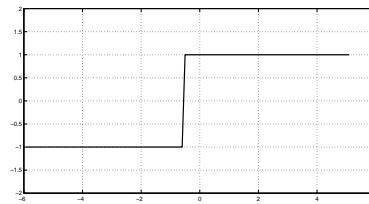


Figura 4.1: Representação de um neurônio artificial.

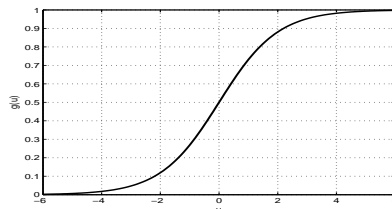
Na Figura 4.1, $x_1; x_2; \dots; x_n$ são os sinais discretos da entrada, $w_1; w_2; \dots; w_n$ são os pesos sinápticos do neurônio, que são utilizados para armazenar o conhecimento adquirido na etapa de aprendizagem a ser descrita na seção 4.5, μ é o limiar de ativação do neurônio que define se a saída será disparada, u é a saída do combinador linear de modo que $u = \sum_{i=1}^N w_i * x_i - \mu$, $g(\cdot)$ é a função de ativação da saída do combinador linear que processa o conjunto de entradas recebidas e as transforma em estado de ativação. A Figura 4. apresenta as principais funções $g(\cdot)$ de ativação usadas em RNAs, destacando-se como as mais usadas a função sigmóide, também conhecida como logística, e a hiperbólica.



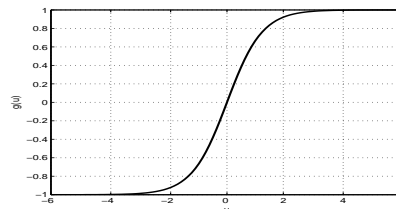
(a) Função rampa.



(b) Função degrau bipolar.



(c) Função sigmóide ou logística.



(d) Função hiperbólica.

Figura 4. : Principais funções de ativação utilizadas em neurônios artificiais.

A função logística é utilizada quando deseja-se que os valores de saída do neurônio estejam no intervalo $[0;1]$, e a função hiperbólica quando se deseja que os valores de saída do neurônio estejam no intervalo $[-1;1]$. A função sigmóide é descrita na Equação 4.1.

$$g(u) = \frac{1}{1 + e^{-\beta u}} \quad (4.1)$$

A função hiperbólica é definida por:

$$g(u) = \frac{1 - e^{\beta u}}{1 + e^{\beta u}} \quad (4.2)$$

Nas equações 4.1 e 4.2 define a inclinação da sigmóide e da hiperbólica. A função rampa é definida por:

$$g(u) = \begin{cases} a; & \text{se } u > a \\ u; & \text{se } -a \leq u \leq a \\ -a; & \text{se } u < a \end{cases} \quad (4.3)$$

onde a denota o valor de ativação mínimo e máximo da função. A função degrau é definida como

$$g(u) = \begin{cases} 1; & \text{se } u \geq 0 \\ 0; & \text{se } u < 0 \end{cases} \quad (4.4)$$

O parâmetro β das funções sigmóide e hiperbólica pode ser ajustado de modo a definir a inclinação da função, como é ilustrado na Figura 4.3.

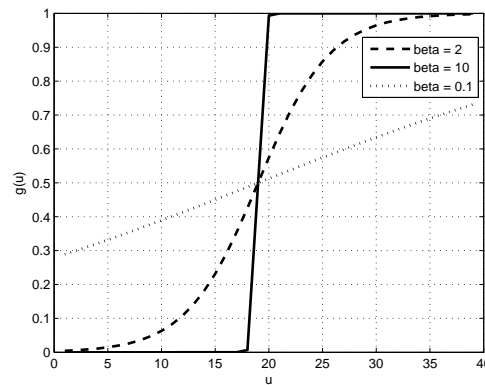


Figura 4.3: Função sigmóide com diferentes valores de β (beta).

Para um γ de valor grande, a função tende a aproximar o comportamento da função degrau, e com um γ pequeno a função tende a aproximar o comportamento da função rampa, ou seja, a partir da função sigmóide ou hiperbólica é possível aproximar outras funções de ativação.

4.3 Arquiteturas de Redes Neurais Artificiais

A forma como os neurônios são agrupados é conhecida como arquitetura de rede neural. Os neurônios artificiais podem ser organizados e conectados de várias formas, resultando em diferentes arquiteturas neurais, com características e aplicações bastante distintas (Grossberg, 1988).

Inicialmente as arquiteturas podem ser classificadas em duas classes, camada única e multi-camadas (Dayhoff, 1990). Nas redes com arquitetura de camada única, tem-se a entrada de dados e uma única camada de neurônios em paralelo que é a própria camada de saída. Um problema relacionado a esse tipo de rede é a sua limitação em representar funções linearmente separáveis, restrição solucionada com a utilização de redes com mais de uma camada, redes conhecidas como multi-camadas. As arquiteturas de redes neurais também podem ser classificadas como recorrentes e não recorrentes (Russell e Norvig, 1995). Esses dois tipos de arquiteturas serão apresentados nas seções 4.3.1 e 4.4 respectivamente.

4.3.1 Redes Neurais Artificiais Não Recorrentes

Redes não recorrentes, também conhecidas como redes multi-camada, são redes que não possuem realimentação de suas saídas para suas entradas (O. Nerrand, 1993). A estrutura das redes neurais não recorrentes pode ser formada por uma camada única ou por múltiplas camadas também conhecidas como redes multi-camadas. As RNAs multi-camadas não recorrentes contém um conjunto de entradas, uma ou mais camadas escondidas e uma camada de saída conforme ilustra a Figura 4.4.

A camada de saída contém os neurônios que fornecem a saída da rede. As camadas que não possuem ligações diretas com a entrada, nem com a saída são denominadas de camadas ocultas. No caso de redes não-recorrentes como na Figura 4.4, não existem conexões ligando um neurônio de uma camada a outro de uma camada anterior, nem a um neurônio da mesma camada, as conexões são feitas somente para neurônios da camada seguinte.

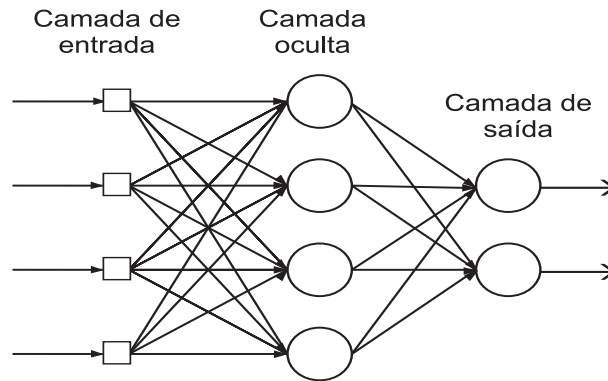


Figura 4.4: Rede neural artificial multi-camada não recorrente.

4.4 Redes Neurais Artificiais Recorrentes para Sistemas Variantes no Tempo

Redes neurais não-recorrentes não são capazes de reutilizar as informações prévias, produzindo apenas mapeamentos estáticos, onde a saída é uma resposta imediata ao estímulo da entrada. Esta é a razão pela qual este tipo de rede neural encontra dificuldade em representar sistemas variantes no tempo, já que o vetor de saída da rede neural depende apenas do vetor de entrada definido no mesmo instante t (Kolen, 1994). Quando deseja-se que todas as informações prévias influenciem no instante de tempo atual t , é recomendável o uso de RNAs do tipo recorrente.

Redes neurais recorrentes são redes que contêm realimentação das saídas para as entradas, sendo suas saídas determinadas pelas entradas atuais e pelas saídas anteriores (Kolen, 1994). As redes recorrentes, quando organizadas em camadas, possuem interligações entre neurônios da mesma camada e entre camadas não consecutivas, gerando interconexões bem mais complexas que as redes neurais não recorrentes (Jain, 001).

A arquitetura mais simples de rede recorrente é a TDNN (*Time-Delay Neural Network*). A TDNN é uma rede perceptron multicamadas (MLP), cujas saídas são re-aplicadas ao longo do tempo, ou seja, a estimação de $x(t+1)$ e $x(t)$ por exemplo, é realizada baseando-se nos valores assumidos por x nos instantes de tempo passados $(t-1)$, $(t-2)$, ..., $(t-n)$. A realimentação dos instantes de tempo $(t-1)$, $(t-2)$, ..., $(t-n)$ da camada de saída para a camada de entrada, permite a criação de representações internas com maior capacidade de memória (Kolen, 1994). A recorrência utilizando valores em instantes de tempo passados, armazena indiretamente os valores

assumidos por x no tempo, configurando-se como uma RNA de maior potencial de modelagem de dados temporais (Zhang e Morris, 1995; Williams e Zipser, 1989). Segundo O. Nerrand (1993), como estruturas de processamento de sinais, redes neurais recorrentes se assemelham a filtros não-lineares com resposta ao impulso infinita. O treinamento da rede TDNN pode ser feito utilizando-se o algoritmo *backpropagation*.

4.4.1 Rede Neural Artificial de Elman

Elman (1990) propôs uma RNA recorrente, adicionando uma camada escondida, denominada camada de contexto. A camada de contexto é responsável por guardar a informação temporal relativa as entradas anteriores da rede. O peso w entre a camada escondida e a camada de contexto é fixado em 1, possibilitando a utilização do algoritmo *backpropagation*. A estrutura da rede de Elman é mostrada na Figura 4.5.

Figura 4.5: Rede Neural Artificial de Elman.

Essa rede diferencia-se das demais soluções temporais por conseguir com que dados de todas as amostras anteriores ao instante de tempo t influenciem no instante presente. Isso é possível devido à realimentação feita na camada escondida, que faz com que todas as amostras anteriores influenciem na saída. Porém, quanto mais recente a amostra maior a influência no resultado.

Kremer (1995) demonstrou que apesar da simplicidade estrutural da rede de Elman, ela é suficientemente capaz de modelar informações temporais utilizando o modelo de neurônio de

Mcculloch-Pitts e o algoritmo de treinamento *backpropagation*.

4.5 Algoritmos de Treinamento de Redes Neurais Artificiais

O treinamento de uma RNA consiste no ajuste dos pesos sinápticos w_j e dos limiares μ . O objetivo no treinamento é fazer com que a rede seja capaz de generalizar um problema, ou seja, não é necessário que a rede conheça todas as possibilidades de entrada para obter a saída, de forma que a aplicação de um conjunto de entradas desconhecidas ao treinamento produza um conjunto de saídas desejadas (Leonard e Kramer, 1990).

Os algoritmos de treinamento de RNAs podem ser classificados em supervisionados ou não supervisionados (Walczak et al., 003). No treinamento supervisionado a rede recebe os vetores de entrada assim como os resultados desejados na saída da rede. A tarefa do treinamento da rede é ajustar os pesos entre os neurônios de forma que a saída seja a desejada, ou seja, a rede é treinada para fornecer a saída desejada a um estímulo de entrada específico. No treinamento de redes multi-camadas semelhantes a Figura 4.4 o algoritmo *backpropagation* também conhecido como retropropagação é o mais citado e utilizado na literatura de RNAs.

De maneira geral, o algoritmo *backpropagation* consiste em dois passos. No primeiro passo, o passo da propagação, a entrada é fornecida aos neurônios da camada de entrada da RNA, e seu sinal se propaga no sentido das conexões (para a frente) em cada uma das camadas da rede até a produção de uma saída na última camada. No passo de propagação, não ocorrem alterações nos pesos sinápticos da rede. O objetivo é apenas calcular a diferença entre a resposta desejada e a saída produzida conforme a equação 4.5.

$$E = \frac{1}{2} \sum_n \sum_{i=1}^p (x_i - \hat{x}_i)^2 \quad (4.5)$$

onde E é a função da medida do erro total da rede, n é o número de padrões apresentados à rede, p é o número de neurônios de saída, x_i é a i -ésima saída desejada e \hat{x}_i é a i -ésima saída gerada pela rede.

No segundo passo, o de retropropagação, ocorre o ajuste dos pesos sinápticos da RNA de modo a minimizar a função E . O ajuste dos pesos deve ser feito proporcionalmente ao negativo do gradiente do erro em relação aos pesos da rede conforme a equação 4.6.

$$\Delta w_{ij} = -\eta * \left(\frac{\partial E}{\partial w_{ij}} \right) \quad (4.6)$$

Onde os índices i e j referem-se ao neurônio i da camada posterior e ao neurônio j da camada anterior e η é a taxa de aprendizado que determina a “velocidade” do passo em direção ao mínimo da função de erro E . Vale ressaltar que um η muito grande poderá levar o treinamento a oscilar em torno do mínimo de E sem conseguir atingi-lo, entretanto um η muito pequeno poderá deixar a convergência do treinamento muito lenta. Uma alternativa é o uso da taxa de aprendizado adaptativa com um η grande no início do treinamento que é reduzido a medida que o erro também diminui.

Calculando-se a derivada para todas as camadas da rede, obtém-se as seguintes fórmulas para a atualização dos pesos w_{ij} , para cada padrão de entrada individualmente.

$$\Delta W_{ij} = \eta * \pm * x_j \quad (4.7)$$

onde

$$\pm = g'_j \left(\sum_{k=1}^{N2} \pm_k w_{kj} \right) \quad (4.8)$$

e g'_j é a derivada da função de ativação do neurônio j .

Para uma apresentação mais detalhada de como o algoritmo *backpropagation* funciona recomenda-se a leitura do artigo original de Rumelhart (Rumelhart e McClelland, 1986).

Apesar de ser o mais utilizado, o algoritmo *backpropagation* apresenta algumas características indesejáveis como por exemplo a baixa velocidade de convergência quando próximo de um mínimo da função de erro E , devido ao fato da atualização dos pesos ser proporcional ao valor do gradiente de erro e problema com mínimos locais, pois o algoritmo geralmente não consegue atingir o mínimo global de E quando atinge um mínimo local.

4.5.1 Algoritmo *Resilient Propagation* (RPROP)

O algoritmo de treinamento *Resilient Propagation* (RPROP) foi proposto por Riedmiller e Braun (1993) como uma alternativa mais eficiente que o *backpropagation* para o treinamento de RNAs do tipo MLP. No algoritmo de treinamento RPROP a alteração dos pesos sinápticos

é baseada em informações de gradiente locais. Para isso, é introduzido, para cada peso, o seu valor individual de atualização D_{ji} , que determina apenas o tamanho da alteração no respectivo peso. Esse valor individual de atualização D_{ji} evolui durante o processo de aprendizado, baseado apenas no sinal local da função de erro E , baseado nas seguintes regras de aprendizado:

$$\Delta_{ji}(n+1) = \begin{cases} \hat{+}\Delta_{ji}(n); \text{ se } \frac{\partial E}{\partial w_{ji}} n \frac{\partial E}{\partial j_i}(n+1) > 0 \\ \hat{-}\Delta_{ji}(n); \text{ se } \frac{\partial E}{\partial w_{ji}} n \frac{\partial E}{\partial j_i}(n+1) < 0 \\ \Delta_{ji}(n); \text{ caso contrário} \end{cases} \quad (4.9)$$

onde $0 < \hat{-} < 1 < \hat{+}$

Uma melhoria sobre a técnica de aprendizagem por retropropagação de erro é a do algoritmo de propagação resiliente (Anastasiadis et al., 2004; Riedmiller e Braun, 1993).

A regra de aprendizado da equação 4.9 faz com que a mudança no sinal das derivadas parciais correspondentes ao peso w_{ij} indica que a última mudança foi grande o suficiente “pulou” sobre um mínimo local da função E , implicando em uma diminuição do valor de Δ_{ji} proporcional ao fator $\hat{-}$. Se a derivada parcial não trocou de sinal, o algoritmo está movendo permanentemente em uma única direção, o que implica num aumento sensível de Δ_{ji} proporcional ao fator $\hat{+}$, para que a convergência no sentido de regiões de menor erro seja acelerada. Uma vez que os valores individuais de atualização Δ_{ji} para cada peso sináptico foram determinados, a atualização dos pesos segue a equação 4.10.

$$\Delta w_{ji}(n+1) = \begin{cases} -\Delta_{ji}(n+1); \text{ se } \frac{\partial E}{\partial w_{ji}}(n+1) > 0 \\ +\Delta_{ji}(n+1); \text{ se } \frac{\partial E}{\partial w_{ji}}(n+1) < 0 \\ 0; \text{ caso contrário} \end{cases} \quad (4.10)$$

É importante citar que o ajuste da taxa de aprendizado e dos pesos sinápticos dependem dos termos do gradiente e não da sua magnitude como no *Backpropagation*. Com a utilização do RPROP, o tamanho da atualização dos pesos depende apenas da sequência de troca de sinais das derivadas parciais.

4.6 Considerações Finais

Neste Capítulo foram apresentadas diferentes arquiteturas de RNAs. As RNAs constituem-se em uma alternativa bastante interessante em relação aos tradicionais modelos matemáticos de predição como o ARIMA. Contudo, as RNAs de maior interesse para esse trabalho são as que consideram o histórico de dados, em particular, destaca-se a RNA de Elman, que é capaz de armazenar indiretamente todo o histórico de dados apresentado a rede.

Capítulo 5

Metodologia

5.1 Considerações Iniciais

Neste Capítulo é apresentada a metodologia proposta na predição das séries temporais econômicas e as medidas de avaliação dos resultados da predição.

5.2 Visão Geral da Técnica

Como já mencionado no Capítulo 1, o objetivo deste trabalho é realizar a predição de séries temporais econômicas. A Figura 5.1 apresenta o resumo da técnica proposta.

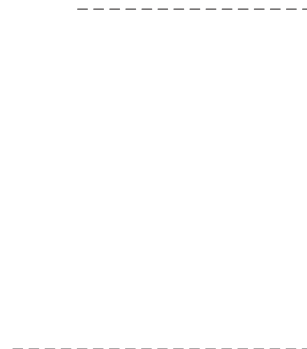


Figura 5.1: Visão geral da técnica proposta.

A predição é determinada a partir de dois módulos principais: um baseado no modelo técnico, que utiliza o histórico das séries temporais econômicas e outro baseado no modelo fundamentalista fazendo uso de variáveis econômicas que tem impacto direto ou indiretos na série. Os dois modelos serão apresentados de forma mais detalhada nas seções 5.3 e 5.4.

5.3 Modelo Técnico

No modelo técnico, os dados sobre as séries temporais econômicas utilizadas foram extraídas diretamente do *website* da BOVESPA¹ (Bolsa de Valores de São Paulo), principal bolsa de valores do Brasil.

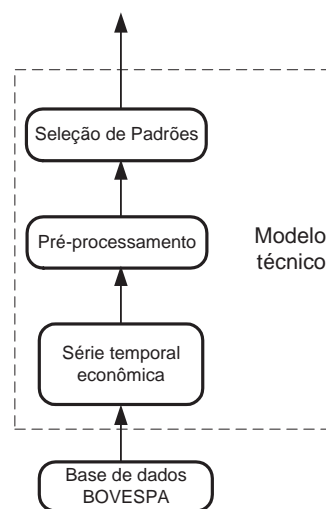


Figura 5. : Diagrama esquemático do modelo técnico.

Conforme o diagrama da Figura 5. , a primeira etapa consiste da obtenção das séries temporais econômicas a partir da base de dados da BOVESPA. No primeiro bloco, as séries foram extraídas isoladamente para cada companhia analisada, de modo que o modelo de série temporal fosse univariado, ou seja, as séries foram analisadas de forma independente uma das outras. Dentre as diversas informações disponibilizadas, no segundo bloco do diagrama são extraídas as informações da série referente a cotação de preço da abertura, preço máximo, preço mínimo, preço médio, preço de fechamento e volume negociado amostrado de forma diária, conforme sugerido nos trabalhos de Yao et al. (1999), Jang (1991) e Lam (2004). Essas variáveis serão

¹<http://www.bovespa.com.br>

referenciadas como variáveis do modelo técnico (VMT) e a predição refere-se ao preço médio dos dias futuros.

Utilizando o *toolbox* de transformada *wavelet* do software *Matlab*, os dados das variáveis VMT são filtrados em uma fase chamada de pré-processamento, conforme ilustrado no terceiro bloco do diagrama do modelo técnico (Diagrama da Figura 5.). Esta etapa objetiva suavizar a componente aleatória da série, ou seja, suavizar o efeito das altas frequências presentes nos dados das séries temporais utilizadas. Segundo a teoria de Down, em séries temporais econômicas, as altas frequências são de difícil compreensão e modelagem, pois estão relacionadas a especulações e boatos, e seus movimentos não possuem nenhuma lógica econômica ou temporal.

O quarto bloco ilustra o emprego de seleção de padrões de treinamento, que visa melhorar o processo de aprendizagem do método de predição e por conseguinte sua precisão em exemplos de teste (Raymer et al., 000; Niska et al., 006).

No contexto de seleção de padrões de treinamento, é possível definir dois tipos de abordagens: métodos que fazem uso de transformações lineares das características a serem selecionadas e métodos que utilizam o espaço de características originais. Em ambos os casos, para cada conjunto original de padrões de entrada objetiva-se obter um conjunto de padrões de dimensão menor. Como exemplo de métodos que fazem uso de transformações lineares é possível citar a função discriminante de Fisher e a análise de componente principais. Algoritmos genéticos e algoritmo de projeções sucessivas são exemplos de métodos que utilizam o espaço de características originais.

Para este trabalho foi escolhido o algoritmo de projeções sucessivas para a seleção dos padrões de treinamento da série temporal econômica, uma vez que este algoritmo mostrou-se promissor quando comparado aos outros métodos citados em problemas de regressão e classificação (Araújo et al., 001; Pontes et al., 005).

5.3.1 Algoritmo de Projeções Sucessivas

O algoritmo de projeções sucessivas (APS) foi proposto inicialmente por Galvão et al. (1999) e Araújo et al. (001), como uma técnica de seleção de variáveis para minimizar problemas de colinearidade em problemas de calibração multivariada. Mais recentemente, no trabalho de Pontes et al. (005), o APS foi empregado também como uma nova estratégia de seleção de variáveis que são minimamente redundantes, mas representativas do conjunto de dados em

problemas de classificação. O uso do APS em problemas de regressão não linear como a predição de série temporais econômicas, constitui-se em uma novidade nas aplicações do APS.

O algoritmo segue um procedimento recursivo em que, partindo de uma variável inicial da matriz de padrões de treinamento \mathbf{X} , é selecionada uma nova variável de maior projeção a cada iteração.

Seja S o conjunto de variáveis disponíveis e que ainda não foram selecionadas, tal que,

$$S = \{\diamond \quad \text{tal que} \quad 0 \leq \diamond \leq J - 1\} \text{ e } j \in \{1, \dots, n(n-1)\} \quad (5.1)$$

onde J é o número total de variáveis disponíveis de \mathbf{X} e \diamond é o número de variáveis que se deseja selecionar.

Seja \mathbf{x}_i a i -ésima e \mathbf{x}_j a j -ésima coluna de \mathbf{X} , tal que \mathbf{X} possua J vetores coluna disponíveis, a projeção de \mathbf{x}_j na direção de \mathbf{x}_i é dada por:

$$\text{Proj}_{\mathbf{x}_i, \mathbf{x}_j} = (\mathbf{x}_i \mathbf{x}_i^T) \mathbf{x}_j (\mathbf{x}_i \mathbf{x}_i^T)^{-1} \quad \forall j \in S \quad (5.2)$$

e a projeção de \mathbf{x}_j no subespaço ortogonal de \mathbf{x}_i é dada por

$$\text{Proj}_{\mathbf{x}_i^\perp, \mathbf{x}_j} = \mathbf{x}_j - \text{Proj}_{\mathbf{x}_i, \mathbf{x}_j} = \mathbf{x}_j - (\mathbf{x}_i \mathbf{x}_i^T) \mathbf{x}_j (\mathbf{x}_i \mathbf{x}_i^T)^{-1} \quad \forall j \in S \quad (5.3)$$

Um exemplo ilustrativo da seleção de variáveis baseada nas projeções sucessivas, extraído de Araújo et al. (2001) e Galvão et al. (1999), está representado na figura 5.3.

Na figura 5.3, é considerada uma matriz \mathbf{X} com 3 amostras e 5 variáveis, que no contexto deste trabalho são os padrões de treinamento. Deseja-se selecionar 3 variáveis, de forma que cada linha de \mathbf{X} seja representada por um vetor em um espaço de 3 dimensões. Tomando-se \mathbf{x}_3 como variável de partida, faz-se $\mathbf{v}_1 = \mathbf{x}_3$. Como é possível observar na figura 5.3, \mathbf{x}_1 tem a maior projeção no subespaço ortogonal a \mathbf{v}_1 . Portanto, faz-se $\mathbf{v}_2 = \mathbf{x}_1$. Para selecionar a terceira variável, repetem-se as projeções, agora sobre o subespaço ortogonal a \mathbf{v}_2 . Como resultado, o vetor \mathbf{x}_5 é escolhido, fazendo-se $\mathbf{v}_3 = \mathbf{x}_5$. Selecionados $\mathbf{x}_1, \mathbf{x}_3$ e \mathbf{x}_5 a escolha é encerrada. Vale ressaltar que segundo Galvão et al. (1999), não seria possível escolher uma quarta variável por esse método, pois todas as variáveis resultantes de uma nova projeção teriam norma zero. Assim, o algoritmo só permite selecionar $n - 1$ variáveis, sendo que n é o número de amostras dos padrões

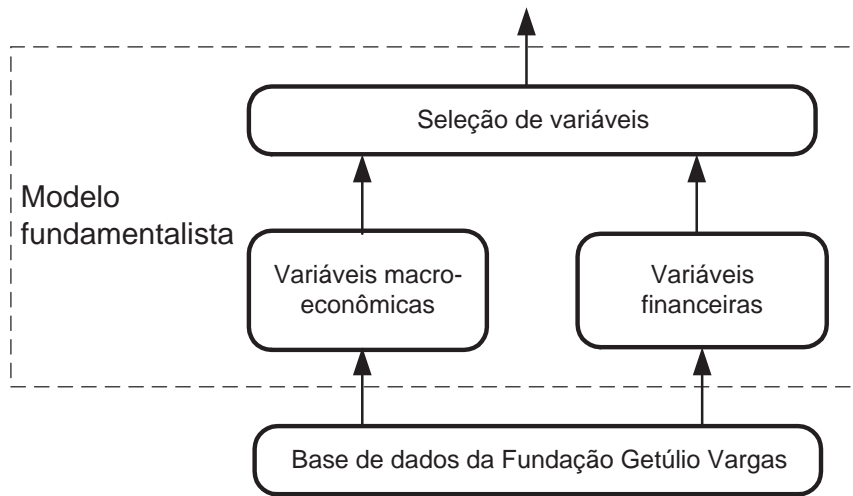


Figura 5.4: Diagrama esquemático do modelo fundamentalista.

manter o conceito relacionado a cada variável. As informações sobre variáveis fundamentalistas foram extraídas a partir do *website* da fundação Getúlio Vargas² e não possuem caráter oficial.

As denominadas variáveis macro-econômicas são variáveis que estão relacionadas ao desempenho geral da economia de mercado nacional ou internacional em que a companhia referente à ação está inserida. Já as variáveis financeiras são variáveis informativas quanto ao desempenho econômico individual da companhia referente a ação, como por exemplo, lucro líquido e dívida ativa (Lameira, 2003; Sá, 1997). O anexo B apresenta uma descrição mais detalhada sobre as variáveis fundamentalistas utilizadas. Essas variáveis serão referenciadas como variáveis do modelo fundamentalista (VMF) e as definições das variáveis utilizadas estão descritas na tabela 5.1. Foram selecionadas cinco variáveis financeiras e seis variáveis macro-econômicas baseando-se nos estudos econômicos de Dhatt (Dhatt et al., 1999b), Nagano (Nagano et al., 2003), Cerqueti (Cerqueti e Rotundo, 2003) e Grôppo (Grôppo, 2004).

A etapa de seleção de variáveis ilustrada na Figura 5.4, consiste no uso do coeficiente de correlação de Spearman (SRCC - *Spearman Rank Correlation Coefficient*) para selecionar as variáveis fundamentalistas que podem estar estatisticamente correlacionadas com a série a ser predita. O coeficiente de Spearman é dado pela equação 5.4.

$$|r| = 1 - \frac{6 \sum_{i=0}^n d_i^2}{n(n^2 - 1)} \quad (5.4)$$

²<http://fgvdados.fgv.br>

Tabela 5.1: Variáveis fundamentalistas.

Variáveis financeiras	
Variável	Descrição
v_1	Total de bens/Endividamento
v_2	Valor de mercado/Total de faturamento
v_3	valor do mercado (Somatório do valor das ações emitidas)
v_4	Lucro líquido
v_5	Mobilidade de capital
Variáveis Macro-econômicas	
v_6	Taxa de juros básica
v_7	Índice de preço ao consumidor
v_8	Cotação do dólar
v_9	Cotação do preço de petróleo internacional
v_{10}	Índice B VESPA [†]
v_{11}	Classificação do risco país [‡]

[†] índice B VESPA indica o desempenho das principais ações negociadas na B VESPA. [‡] Risco país é um método mundialmente reconhecido que classifica o risco de calote da dívida pública de um país e conseqüentemente o risco de se investir no país.

onde n é o número de pares entre as variáveis x_1 e x_2 a serem analisadas e d_i é expresso na equação 5.5.

$$d_i = (\text{Posto de } x_{1i} \text{ dentre os valores de } x_1) - (\text{Posto de } x_{2i} \text{ dentre os valores de } x_2) \quad (5.5)$$

O coeficiente de Pearson é um teste estatístico freqüentemente utilizado para encontrar a correlação linear entre variáveis. Entretanto, segundo Jaeger (1990), o teste de correlação de Spearman além de não requerer a suposição de linearidade, como no coeficiente de Pearson, apresenta duas características importantes em comparação ao coeficiente de Pearson. A primeira característica é a possibilidade de se fazer correlações de dados de fontes distintas e grandezas diferentes. A correlação de Spearman considera tendências e não valores como a correlação de Pearson, ou seja, Spearman equivale a um teste qualitativo de correlação e Pearson a um teste quantitativo (Jaeger, 1990; Osborn, 2006). A segunda característica importante é que o coeficiente de Spearman não requer nenhuma suposição sobre a distribuição de freqüências das variáveis, ao contrário do coeficiente de Pearson que requer que as variáveis analisadas estejam normalmente distribuídas.

5.5 Predição: Combinando Modelo Técnico e Fundamentalista

Após a conversão e filtragem das variáveis VMT e da seleção das variáveis VMF para cada série estudada, a rede neural de Elman foi elaborada de modo a prever p valores futuros. Utilizou-se os dados de modelo técnico e fundamentalista do período de janeiro de 1999 a agosto de 2006, para treinamento da rede neural de Elman descrita na seção 4.4, sendo que os dados do modelo técnico de agosto a novembro de 2006 foram utilizados para validar a predição. Vale salientar

$$RMSEP = \sqrt{\frac{1}{N} \sum_{t=1}^N (x(t) - \hat{x}(t))^2} \quad (5.6)$$

onde $x(t)$ é o valor real da série no tempo t , $\hat{x}(t)$ é o valor predito no tempo t e N é o número de observações no conjunto de validação.

Para cada série analisada, são executadas T treinamentos e predições onde são calculadas medidas estatísticas conforme as Equações 5.7 e 5.8.

$${}^1_{rmsep} = \frac{\sum_{i=0}^T RMSEP_i}{T} \quad (5.7)$$

$$\mathcal{Y}_{rmsep} = \sqrt{\frac{\sum_{i=0}^T (RMSEP_i - {}^1_{rmsep})^2}{T}} \quad (5.8)$$

O parâmetro ${}^1_{rmsep}$ calcula a média dos erros obtidos com T treinamentos e predições de uma mesma série temporal, enquanto que o parâmetro \mathcal{Y}_{rmsep} calcula o desvio padrão das T medidas de erro $RMSEP$.

5.6 Software de Desenvolvimento

Para a implementação da técnica, foi utilizado o software MATLAB (*Matrix Laboratory*) na versão 7.0, com os *toolbox* de redes neurais artificiais e transformada *wavelet*. O anexo C apresenta o software desenvolvido.

A leitura dos dados da BOVESPA foi realizada por meio de um algoritmo desenvolvido na linguagem JAVA, na versão 5.0, com o objetivo de ler o arquivo binário e convertê-lo em um formato de fácil leitura no software MATLAB.

5.7 Considerações Finais

Este Capítulo apresentou a técnica proposta para o problema de predição de séries temporais econômicas. A técnica é composta por dois módulos principais, sendo um baseado no modelo técnico e outro no fundamentalista. A combinação desses dois modelos pode apresentar melhores resultados em relação as técnicas que fazem uso de apenas um dos modelos. Apresentou-se as

medidas de avaliação de erro de predição por meio das equações 5.6, 5.7 e 5.8.

Capítulo 6

Resultados

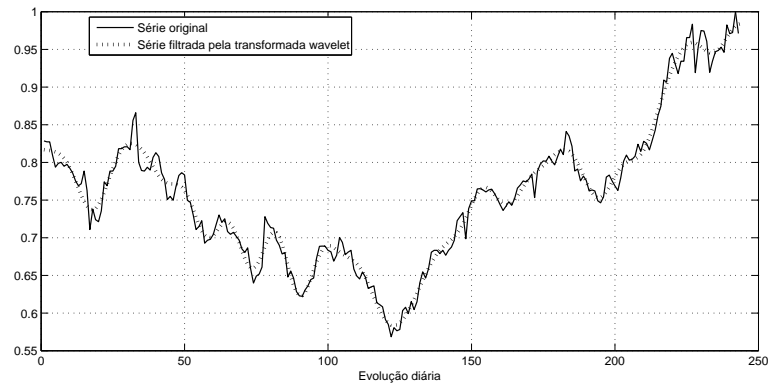
6.1 Considerações Iniciais

Neste Capítulo são apresentados e discutidos os resultados obtidos com o método proposto para a predição de séries temporais econômicas. Deste modo, objetiva-se efetuar uma análise da filtragem de comportamentos aleatórios presentes na série, da seleção de padrões por meio do algoritmo APS e avaliar a importância de variáveis fundamentalista na entrada da rede neural artificial. Os resultados são comparados com as técnicas clássicas apresentadas no Capítulo e com modelos de predição baseados em análise técnica.

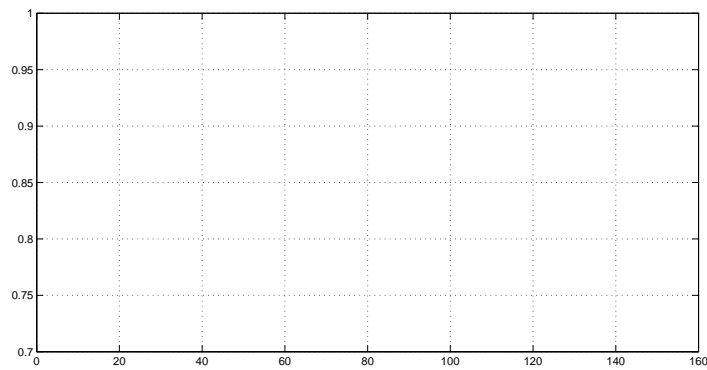
6.2 Suavização da Componente Aleatória das Séries

Conforme descrito no Capítulo 3, aplica-se a transformada *wavelet*, sendo que para cada canal h aplica-se o limiar universal (Equação 3.14) em cada nível de decomposição. Experimentalmente definiu-se que três níveis de decomposição são suficientes para suavizar a componente aleatória. Para a maioria das séries estudadas, a filtragem utilizando níveis maiores que três, distorce parte do comportamento da série. A Figura 6.1 ilustra dois exemplos de suavização.

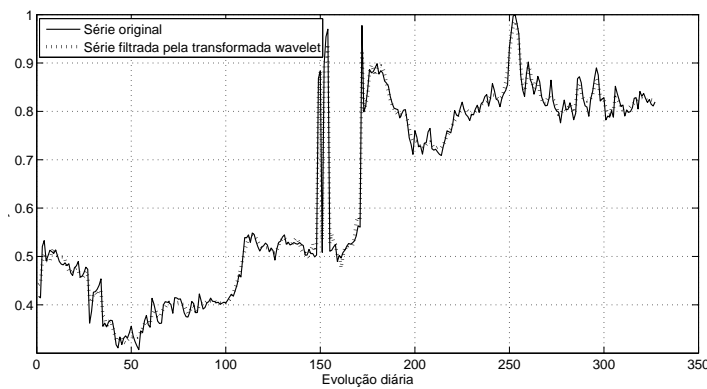
Na Figura 6.1 são apresentados trechos das séries históricas de cotações de preço médio no período de agosto a dezembro de 2005. Note que houve uma suavização das altas frequências das séries, sendo que o efeito da suavização lembra um alisamento polinomial. Entretanto, em séries onde os picos são altos e estreitos, como a série da Figura 6.1(c), a transformada *wavelet* também consegue suavizar a série.



(a) Filtragem da série econômica das ações do Grupo Pão de açúcar.



(b) Filtragem da série econômica das ações do grupo ITAUSA.



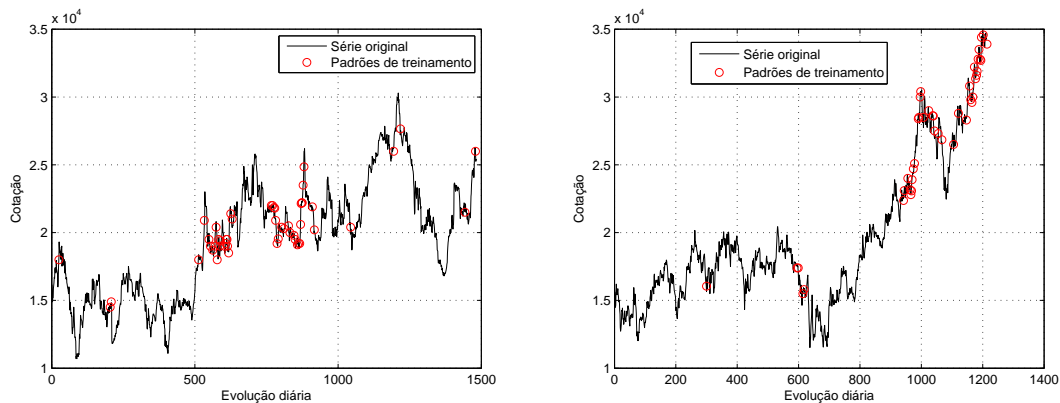
(c) Filtragem da série econômica das ações da Embraer.

Figura 6.1: Filtragem das séries utilizando transformada *wavelet*.

É possível observar na Figura 6.1(c) que picos em intervalos de tempo curtos com baixas frequências não são cortados na filtragem, fato que não é possível utilizando transformada de Fourier ou outra técnica de alisamento. Os trechos específicos contendo alta frequência são suavizados, como por exemplo no intervalo de tempo de 0 a 100 da Figura 6.1(c), preservando picos com baixa frequência, como por exemplo no intervalo de tempo de 130 a 170 da Figura 6.1(c). Isto é possível devido à capacidade de resolução tempo-frequência da transformada *wavelet* descrita no Capítulo 3.

6.3 Seleção de Padrões Utilizando o APS

Como relatado na seção 5.3.1, o APS foi adaptado para o problema de predição de série temporais econômicas para selecionar os padrões de treinamento mais representativo do comportamento da série estudada. Dois resultados da seleção de padrões de treinamento por meio do algoritmo APS em duas séries temporais econômicas diferentes são indicados na Figura 6. .



(a) Série original e padrões selecionados para a Coteminas.

(b) Série original e padrões selecionados para o Itaú.

Figura 6. : Resultado da seleção de padrões (Círculos vermelhos) por meio do algoritmo APS nas séries da Coteminas e Itaú.

No caso da série da Coteminas, o algoritmo selecionou a maioria dos padrões próximos aos instantes de tempo $t = 500$ e $t = 800$. A matriz de padrões continha inicialmente 1357 padrões, sendo que o melhor conjunto selecionado pelo APS em termos de RMSEP contém 50 padrões, proporcionando uma redução da matriz de treinamento em torno de 96%. Já para a série do Itaú, o algoritmo APS selecionou padrões bem próximos aos instantes de tempo a serem preditos. A matriz de padrões continha inicialmente 14 padrões de treinamento, sendo reduzido para 30

com o APS, ou seja, em termos percentuais 97% menor.

6.4 Combinando Modelo Técnico e Fundamentalista

Com a série suavizada aplica-se o teste SRCC da seção 5.4, para adição de variáveis fundamentalistas. O resultado do teste SRCC com arredondamento para duas casas decimais de precisão, e o resumo das variáveis escolhidas para cada série analisada é mostrado na Tabela 6.1.

Tabela 6.1: Resultado do coeficiente de Spearman para as variáveis fundamentalistas.

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	Variáveis selecionadas
C_1	0,19	0,23	0,05	0,20	0,01	0,08	0,12	0,03	0,01	0,13	0,27	$v_1, v_2, v_4, v_7, v_{10}$ e v_{11}
C_2	0,10	0,04	0,08	0,22	0,05	0,13	0,08	0,07	0,01	0,11	0,23	v_1, v_4, v_6, v_{10} e v_{11}
C_3	0,12	0,22	0,05	0,21	0,03	0,01	0,01	0,04	0,01	0,10	0,20	v_1, v_2, v_4, v_{10} e v_{11}
C_4	0,11	0,12	0,01	0,21	0,04	0,01	0,01	0,11	0,02	0,12	0,27	$v_1, v_2, v_4, v_8, v_{10}$ e v_{11}
C_5	0,11	0,01	0,01	0,18	0,00	0,10	0,07	0,09	0,19	0,21	0,19	v_1, v_4, v_9, v_{10} e v_{11}
C_6	0,12	0,14	0,02	0,23	0,03	0,06	0,10	0,03	0,01	0,15	0,23	$v_1, v_2, v_4, v_4, v_{10}$ e v_{11}
C_7	0,00	0,02	0,08	0,2	0,01	0,06	0,10	0,00	0,00	0,1	0,09	v_1, v_2 e v_{11}
C_8	0,00	0,02	0,08	0,2	0,01	0,06	0,10	0,00	0,00	0,1	0,09	v_1, v_2
C_9	0,00	0,02	0,08	0,2	0,01	0,06	0,10	0,00	0,00	0,1	0,09	v_4, v_6, v_7 e v_{11}
C_{10}	0,00	0,00	0,00	0,0	0,01	0,00	0,00	0,00	0,00	0,1	0,00	nenhuma
C_{11}	0,00	0,00	0,07	0,04	0,01	0,02	0,00	0,01	0,00	0,23	0,02	v_{10}
C_{12}	0,00	0,02	0,08	0,20	0,01	0,06	0,10	0,00	0,00	0,21	0,09	v_4 e v_{10}
C_{13}	0,01	0,03	0,16	0,18	0,08	0,06	0,00	0,01	0,00	0,21	0,09	v_{10}
C_{14}	0,21	0,00	0,06	0,20	0,01	0,00	0,21	0,12	0,17	0,03	0,15	v_1, v_7 e v_{11}
C_{15}	0,02	0,07	0,00	0,21	0,11	0,01	0,22	0,01	0,01	0,02	0,08	v_4 e v_7

onde C_1 : Grupo Pão de Açúcar, C_2 : Grupo ITAUSA, C_3 : Embraer, C_4 : Companhia Vale do rio doce, C_5 : Petrobrás, C_6 : AMBEV, C_7 : Natura, C_8 : VIV , C_9 : TELES P, C_{10} : G L, C_{11} : TAM, C_{12} : CPFL Energia, C_{13} : Aracruz, C_{14} : Sadia e C_{15} : Coteminas.

Os resultados da Tabela 6.1 mostram que para cada série econômica diferentes variáveis podem estar correlacionadas com o comportamento da série. Para a série do Pão de Açúcar as variáveis v_2, v_4, v_7, v_{10} e v_{11} podem estar correlacionadas com o comportamento da série e as variáveis v_1, v_3, v_5, v_6, v_8 e v_9 não apresentam contribuição significativa para o comportamento da série do grupo Pão de Açúcar. Entretanto para a série da Petrobrás as variáveis v_1, v_4, v_6, v_{10} e v_{11} podem estar correlacionadas com o comportamento da série e v_2, v_3, v_5, v_6 e v_8 não. Portanto, para cada série econômica, diferentes variáveis podem contribuir para a compreensão do comportamento e consequentemente para sua estimação. Para cada série temporal econômica as variáveis fundamentalistas escolhidas são adicionadas na entrada da rede neural para estimar o comportamento futuro.

Entretanto para algumas séries analisadas o teste estatístico não foi capaz de identificar qualquer relação para a maioria das variáveis analisadas. Para a companhia GOL (C_{10}), por

exemplo, nenhuma variável fundamentalista é adicionada ao conjunto de variáveis VMT, não contribuindo em nada com a modelagem dos dados. É possível que o intervalo de dados utilizados, não seja o mais adequado, ou que para algumas séries, a mudança de indicadores econômicos ocorre com um *delay* de tempo desconhecido, e não de forma imediata como alguns pesquisadores sugerem. Outra possibilidade é a de que as variáveis fundamentalistas exerçam influência em determinados trechos da série temporal econômica, sugerindo que o teste deva ser aplicado em janelas de tempo Δt , ao invés da aplicação do teste de correlação sobre toda a série.

As variáveis fundamentalistas selecionadas compõem o método proposto no trabalho, que ainda faz uso da transformada *wavelet* e do algoritmo APS, e seus resultados são comparados na Tabela 6. com técnicas clássicas como os modelos ARIMA e técnico.

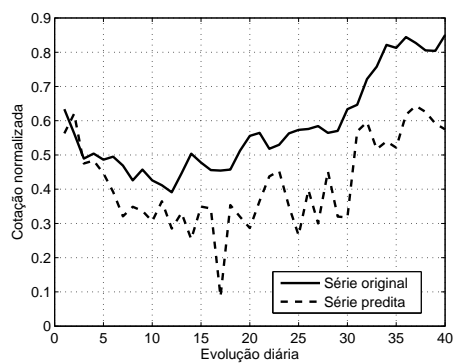
Tabela 6. : Resultados RMSEP com técnicas clássicas e a técnica proposta. ¹Método A: Modelo ARIMA, ²Método B: Modelo técnico, ³Método C: Modelo técnico combinado com o modelo fundamentalista.

	Método A ¹	Método B ²		Método C ³	
	RMSEP	RMSEP		RMSEP	
		$\mu(10^b)$	$\sigma(10^b)$	$\mu(10^b)$	$\sigma(10^b)$
C_1	$5,2 \times 10^{-1}$	$4,1 \times 10^{-2}$	$1,0 \times 10^{-2}$	$1,8 \times 10^{-2}(50\Diamond)$	$2,3 \times 10^{-3}(50\Diamond)$
C_2	$8,6 \times 10^{-2}$	$4,1 \times 10^{-2}$	$1,0 \times 10^{-2}$	$1,8 \times 10^{-2}(30\Diamond)$	$2,3 \times 10^{-3}(30\Diamond)$
C_3	$9,1 \times 10^{-1}$	$5,1 \times 10^{-1}$	$3,7 \times 10^{-3}$	$2,9 \times 10^{-1}(80\Diamond)$	$2,2 \times 10^{-3}(80\Diamond)$
C_4	$1,1 \times 10^{-1}$	$7,9 \times 10^{-2}$	$3,0 \times 10^{-3}$	$5,6 \times 10^{-2}(80\Diamond)$	$1,6 \times 10^{-3}(80\Diamond)$
C_5	$9,8 \times 10^{-1}$	$6,9 \times 10^{-1}$	$2,1 \times 10^{-3}$	$4,0 \times 10^{-1}(119\Diamond)$	$9,4 \times 10^{-4}(119\Diamond)$
C_6	$1,4 \times 10^{-1}$	$7,8 \times 10^{-1}$	$4,4 \times 10^{-3}$	$3,3 \times 10^{-1}(80\Diamond)$	$2,1 \times 10^{-3}(80\Diamond)$
C_7	$3,2 \times 10^{-1}$	$2,6 \times 10^{-1}$	$5,7 \times 10^{-3}$	$1,9 \times 10^{-2}(50\Diamond)$	$4,1 \times 10^{-3}(50\Diamond)$
C_8	$8,2 \times 10^{-2}$	$7,3 \times 10^{-2}$	$2,5 \times 10^{-3}$	$6,8 \times 10^{-2}(80\Diamond)$	$1,0 \times 10^{-3}(119\Diamond)$
C_9	$7,9 \times 10^{-2}$	$5,3 \times 10^{-2}$	$6,1 \times 10^{-3}$	$2,2 \times 10^{-2}(119\Diamond)$	$5,2 \times 10^{-3}(119\Diamond)$
C_{10}	$3,3 \times 10^{-1}$	$1,2 \times 10^{-1}$	$9,9 \times 10^{-3}$	$9,7 \times 10^{-2}(80\Diamond)$	$8,8 \times 10^{-3}(80\Diamond)$
C_{11}	$9,6 \times 10^{-2}$	$4,2 \times 10^{-2}$	$3,8 \times 10^{-3}$	$3,5 \times 10^{-2}(50\Diamond)$	$3,0 \times 10^{-3}(50\Diamond)$
C_{12}	$5,7 \times 10^{-2}$	$3,8 \times 10^{-2}$	$9,3 \times 10^{-3}$	$1,9 \times 10^{-2}(119\Diamond)$	$4,9 \times 10^{-3}(119\Diamond)$
C_{13}	$2,0 \times 10^{-1}$	$3,4 \times 10^{-2}$	$4,2 \times 10^{-3}$	$2,4 \times 10^{-2}(80\Diamond)$	$5,3 \times 10^{-3}(80\Diamond)$
C_{14}	$6,5 \times 10^{-2}$	$2,9 \times 10^{-2}$	$4,3 \times 10^{-3}$	$1,8 \times 10^{-2}(119\Diamond)$	$2,7 \times 10^{-3}(119\Diamond)$
C_{15}	$2,5 \times 10^{-1}$	$3,7 \times 10^{-2}$	$7,4 \times 10^{-3}$	$1,7 \times 10^{-2}(50\Diamond)$	$2,5 \times 10^{-3}(50\Diamond)$

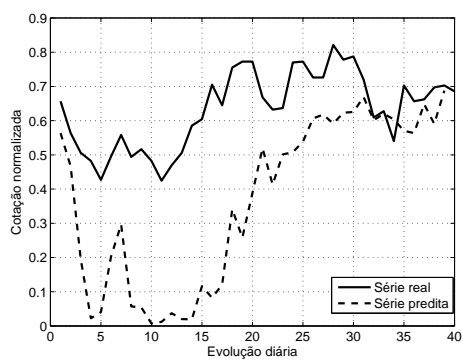
\Diamond : Quantidade de padrões selecionados pelo APS.

b : Quantidade de execuções utilizadas para o cálculo dos parâmetros μ e σ .

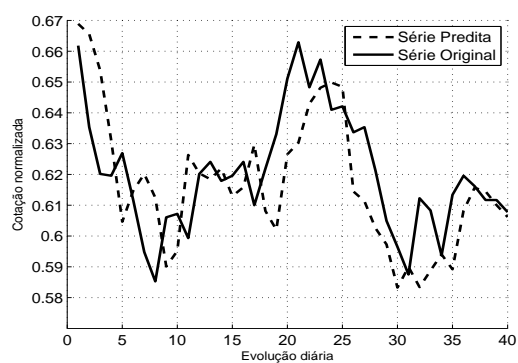
Os resultados apresentados na Tabela 6. mostram que a predição com o modelo ARIMA tem pouca habilidade de predição (Método A), devido às limitações da suposição de linearidades discutidas no Capítulo . Os resultados obtidos com RNA (Método B) são melhores, mas ainda inferiores aos obtidos com o método proposto (Método C). Alguns resultados da predição do modelo ARIMA e do método proposto são apresentados nas Figura 6.3.



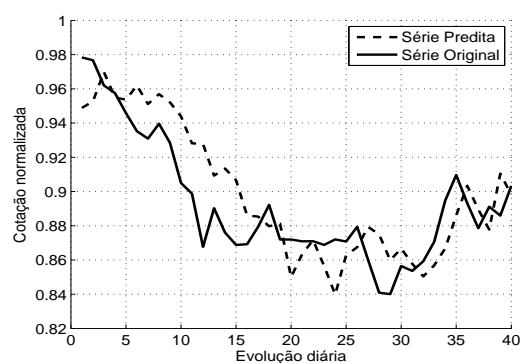
(a) Predição da série econômica do Grupo Pão de Açúcar utilizando o modelo ARIMA.



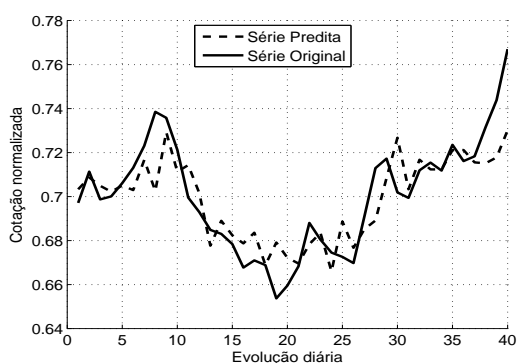
(b) Predição da série econômica do Grupo ITAUSA utilizando o modelo ARIMA.



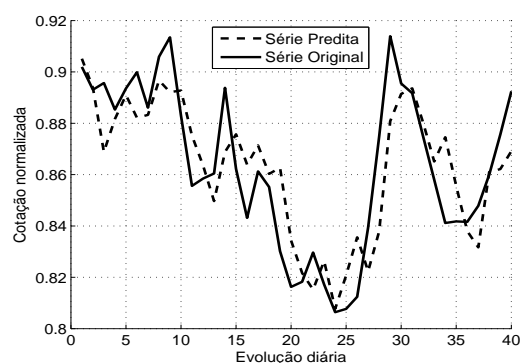
(c) Predição da série econômica da Aracruz utilizando o método proposto.



(d) Predição da série econômica do ITAUSA utilizando o método proposto.



(e) Predição da série econômica do Grupo Pão de Açúcar utilizando o método proposto.



(f) Predição da série econômica da Cemig utilizando o método proposto.

Figura 6.3: Predição de séries econômicas com o método ARIMA e o método proposto.

Finalmente, quando os três métodos são comparados, é possível notar que a adição de variáveis fundamentalistas e seleção de padrões por meio do APS oferece em média uma melhora de 76% em relação ao método ARIMA e 47% em relação ao método clássico baseado no modelo técnico e RNA. Além disso, o uso do APS melhorou substancialmente o tempo de treinamento e predição da RNA, em razão da redução do número de padrões de treinamento. A técnica proposta apresenta resultados pobres nas séries onde poucas variáveis fundamentalista são adicionadas ao modelo de dados utilizado na predição, oferecendo resultados similares ou inferiores ao método B, e pouco superiores ao método A. O que leva a inferir que, de fato, as chamadas informações fundamentalistas, são relevantes no estudo do problema de predição de séries temporais econômicas de cotações de ações de empresas estudadas neste trabalho.

Apesar de ser um modelo bastante utilizado por profissionais da área de economia, o modelo fundamentalista recebeu uma recente atenção da comunidade científica da área de predição de séries. Entretanto, esta nova tendência ainda não havia sido discutida e implementada nas técnicas de inteligência artificial como as redes neurais artificiais.

6.5 Considerações Finais

Neste Capítulo foram apresentados os resultados obtidos com a técnica proposta. Os resultados foram comparados com os procedimentos de predição clássicos encontrados na literatura. Demonstrou-se experimentalmente que a técnica proposta possui uma capacidade de predição melhor que as técnicas tradicionais baseadas nos modelos técnico e ARIMA.

Capítulo 7

Conclusões

Neste trabalho foi apresentado um método de predição não linear de séries temporais econômicas. O método baseia-se no uso de filtro wavelet multiescala, modelos técnicos e fundamentalista de análise de séries temporais econômicas e redes neurais artificiais. As séries obtidas após a decomposição contém informações sobre a tendência da série com o comportamento aleatório suavizado. Os resultados apresentados mostram que a transformada wavelet é um ferramenta eficiente na suavização da componente aleatória de séries temporais econômicas. O uso do algoritmo de projeções sucessivas na seleção de padrões de treinamento, foi importante para reduzir o tempo de treinamento da RNA, uma vez que o conjunto de padrões foi reduzido de forma considerável. Utilizando a RNA de Elman foi possível fazer predições de cotações futuras com a adição de variáveis fundamentalistas que influenciam no comportamento da série. A combinação dos modelos técnico e fundamentalista, caracterizado pela adição das variáveis fundamentalistas na série histórica de preços apresentou um melhor resultado quando comparado com métodos tradicionais que são baseados somente no modelo técnico.

7.1 Contribuições do Trabalho

Na análise dos resultados obtidos foi possível identificar algumas contribuições do trabalho para a solução do problema de predição de séries temporais econômicas. Os resultados apresentados na tabela 6. , indicam que não é correto tratar uma série econômica como uma série temporal pura, ou seja, fazer predições baseando-se somente em seu comportamento histórico. A predições futuras sem variáveis econômicas podem ser feitas, mas quais são as condições econômicas pré-determinadas para estimar o comportamento futuro? Este trabalho demonstra que por

melhor que seja o método de predição, é extremamente relevante analisar os aspectos econômicos relacionados ao problema. Essa abordagem foi inicialmente proposta no recente trabalho de Lam (004). Entretanto, este trabalho apresentou resultados similares aos do trabalho de Lam (004) em termos de erro de predição, porém com menos variáveis fundamentalistas adicionadas às séries, proporcionando um modelo mais robusto do que o proposto por Lam (Lam, 004). Adicionalmente, foi possível constatar que métodos de seleção de variáveis, podem ser aplicados ao problema de seleção de padrões de treinamento. O uso do APS para tal finalidade constituiu-se em uma novidade na gama de aplicações de tal método.

Vale ressaltar também que as séries econômicas de países subdesenvolvidos ou em desenvolvimento são consideradas de “comportamento hostil”, ou seja, de difícil compreensão e predição quando comparadas ao comportamento da maioria das séries econômicas obtidas em países com economia estável.

7.2 Publicações

A partir do estudo desenvolvido nesta dissertação de mestrado, foram elaborados dois artigos científicos, um já publicado na *Lecture Notes in Computer Science* (SOARES, A. S., PAIVA, M. S. V, COELHO, C. J. Technical and Fundamental Analysis for the Forecast of Financial Scrip Quotation: An Approach Employing Artificial Neural Networks and Wavelet Transform. SPRINGER - LECTURE NOTES IN COMPUTER SCIENCE, v.4493, p.104 - 103 , 007.) e outro que na data da defesa desta dissertação encontrava-se em fase final de revisão e submissão com o título provisório “Features Selection Based on Successive Algorithm Projections for Artificial Neural Networks in Financial Time Series Forecasting”.

Os resultados também contribuíram em um capítulo do livro “Redes Neurais Artificiais para Engenheiros e Cientistas”, de autoria do Professor Doutor Ivan Nunes da Silva, professor associado do departamento de Engenharia Elétrica da Escola de Engenharia Elétrica de São Carlos, na Universidade de São Paulo, que encontra-se em fase final de publicação.

7.3 Limitações

Apesar do uso do teste estatístico SCRC para a seleção das variáveis fundamentalistas, o limiar de decisão estabelecido constitui um ponto crítico. Não foi estabelecido um teste de

hipótese para a inclusão ou não da variável.

O uso comercial do método de predição deve levar em consideração critérios como desconto de risco, taxa de mobilidade de capital para se comprar ou vender uma ação e cálculo de incerteza sobre a predição realizada.

7.4 Trabalho Futuros

A partir do desenvolvimento e apresentação dos resultados deste trabalho, foi possível observar algumas sugestões de trabalhos futuros:

- O estudo de novas variáveis fundamentalistas poderá ser continuado.
- Aprimoramento do filtro *wavelet* de suavização da série.
- Uso de técnicas de agrupamento de aprendizagem de máquina como *bagging* e *subbagging*.
- Aprimorar o critério de seleção de variáveis, como por exemplo o uso de um teste estatístico de hipótese nula de que a variável é ou não co-relacionada com a série.
- Aplicar o critério de seleção de variáveis sobre trechos da série e não em toda a série.

Referências Bibliográficas

- Al-Deehani, T. M. (006). Seasonality as an unobservable component: the case of kuwait stock market. *Applied Financial Economics* 16(6), 471–478(8).
- Al-Saad, K. (005). Seasonality in stock returns: evidence from an emerging market. *Applied Financial Economics* 15(1), 63–71.
- Anastasiadis, A. D., G. D. Magoulas, e M. N. Vrahatis (004). New globally convergent training scheme based on the resilient propagation algorithm. *Neurocomputing* 64, 53– 70.
- Araújo, M. C. U., T. C. B. Saldanha, R. K. H. Galvão, T. Yoneyama, H. C. Chame, e V. Visani (001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems* 57(), 65–73.
- Attia, J. O. (1999). *Fourier Analysis: Electronics and Circuit Analysis using MATLAB*. Boca Raton: CRC Press LLC.
- Aussem, A. e F. Murtagh (1997). Combining neural network forecast on wavelet-transformed time series. *Connection Science* 9(1), 113–1 .
- Box, G. e G. Jenkins (1976). Time series analysis, forecasting and control. *Rev. Ed. San Francisco: Holden Day*.
- Box, G., G. M. Jenkins, e G. Reinsel (1976). *Time Series Analysis: Forecasting and Control* (ed.). Prentice Hall.
- Bracewell, R. N. (1988). *Fast Fourier Transform and Its Applications* (1 ed.). Prentice Hall.
- Brown, S. J., W. N. Goetzmann, e A. Kumar (1998). The dow theory: William peter hamilton’s track record reconsidered. *The Journal of Finance* 53(4), 1311–1333.

- Cajueiro, D. O. e B. M. Tabak (005). Possible causes of long-range dependence in the brazilian stock market. *Physica A: Statistical Mechanics and its Applications* 345 Pages 635-645.(3), 635–645.
- Cerqueti, R. e G. Rotundo (003). Microeconomic modeling of financial time series with long term memory. *Proceedings of IEEE International Conference on Computational Intelligence for Financial Engineering*, 191–198.
- Claerbout, J. F. (1999). *Fourier Transforms and Waves: in four lectures*. Stanford University.
- Clarke, B. (000). *Fourier Theory*. Macquarie University.
- Connor, J. T., R. D. Martin, e L. E. Atlas (1994). Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks* 5(), 40.
- Daubechies, I. (1990). The wavelet transform, time-frequency localization and signalanalysis. *IEEE Transaction on Information Theory* 36(5), 961–1005.
- Dayhoff, J. E. (1990). *Neural Network Architectures*. Van Nostrand Reinhold.
- Dhatt, M. S., Y. H. Kim, e S. Mukherji (1999a). Relations between stock returns and fundamental variables: evidence from a segmented market. *Asia-Pacific Financial Markets* 6, 1– 33.
- Dhatt, M. S., Y. H. Kim, e S. Mukherji (1999b). Relations between stock returns and fundamental variables: Evidence from a segmented market. *Asia-Pacific Financial Markets* 6, 1– 33.
- Diniz, H., L. de Andrade, A. de Carvalho, e M. de Andrade (1999). Architecture design of artificial neural networks based on box and jenkins models for time series prediction. *Third Internacional Conference on Computational Intelligence and Multimedia Applications*, 9–34.
- Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory* 41(3), 613–6 7.
- Downie, T. e B. Silverman (1998). The discret multiple wavelet transform and thresholding methods. *IEEE Transactions on Information Theory* 46(9), 558– 561.
- Elliott, D. F. e K. R. Rao (198). *Fast Transforms: Algorithms, Analyses and Applications*. Florida - EUA: Academic Press, INC.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science* 14(1), 179– 1.

-
- Fryzlewicz, P. Z. (003). *Wavelet Techniques for Time Series and Poisson Data*. Tese de doutorado, University of Bristol.
- Galvão, R. K. H., T. C. B. Saldanha, T. Yoneyama, e M. C. U. D. Araújo (1999). Uma técnica para a redução de cardinalidade de padrões de entrada de uma rede neural artificial. *Proceedings of the IV Brazilian Conference on Neural Networks*, 1 4–1 9.
- Gencay, R., B. Whitcher, R. Gengay, e F. Selguk (001). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press.
- Granger, C. W. J. e M. J. Morris (1976). Time series modelling and interpretation. *Journal of the Royal Statistical Society* 139(), 46– 57.
- Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks* 1(1), 17–61.
- Grôppo, G. S. (004). Causabilidade das variáveis macroeconômicas sobre o ibovespa. Masters thesis, Escola Superior de Agricultura Luiz de Queiroz - Universidade de São Paulo, Piracicaba - SP - Brasil.
- Gurgul, H., R. Mestel, e C. Chleicher (003). Stock market reactions to dividend announcements: empirical evidence from the austrian stock market. *Financial markets and portfolio management* 17(3), 33 –350.
- Hamilton, J. D. (1954). *Time Series Analysis*. New Jersey - USA: Princeton University Press.
- Harrison, P. e C. Stevens (1976). Bayesian forecasting. *Journal of The Royal Statistical Society* 38(3), 05– 67. Series B.
- Harvey, A. C. e P. Todd (1983). Forecasting economic time series with structural and box-jenkins models: A case study. *Journal of Business and Economic Statistics* 1(4), 99.
- Haykin, S. (001). *Redes Neurais: Princípios e Práticas* (ed.). Bookman.
- Hernandez, E. e G. Weiss (1996). *A First Course on Wavelets*. CRC Press.
- Hotta, L. K. (1988). Seasonal adjustment of brazilian time series. *Journal of Econometrics* 8(1), 83–95.
- Huang, S.-J. e C.-T. Hsieh (00). Coiflet wavelet transform applied to inspect power system disturbance-generated signals. *IEEE Transactions on Aerospace and Electronic Systems* 38(1), 04– 10.

- Ingle, V. K. e J. G. Proakis (000). *Digital Signal Processing*. Bookware Companion Series. Northeastern University: Brooks/Code Publishing Company.
- Jaeger, R. M. (1990). *Statistics: A Spectator Sport* (Secord Edition ed.). London - U.K: Sage Publications Inc.
- Jain, L. (001). *Recurrent Neural Netoworks: Design and Applications*. Departments of Physics and Computer Science and Information Systems American University Washington, D.C.: CRC Press.
- Jang, G. S. (1991). An intelligent stock portfolio management system based on short-term trend prediction using dual-module neural networks. *International Conference on Artificial Neural Networks 1*, 447–45 .
- Kaiser, G. (1994). *A Friendly Guide to Wavelets*. University of Massachusetts at Lowell - USA: Birkhäuser.
- Kedem, B. e K. Fokianos (00). *Regression Models for Time Series Analysis*. University of Maryland and Cyprus: John Wiley Sons.
- Kolen, J. (1994). *Exploring the Computational Capabilities of Recurrent Neural Networks*. Ph.d. thesis, The Ohio State University, USA.
- Koskela, T., M. Lehtokangas, J. Saarinen, e K. Kaski (1996). Time series prediction with multi-layer perceptron.
- Kremer, S. (1995). On the computational power of elman-style recurrent networks. *IEEE Trans. on Neural Networks 6*(4), 1000–1004.
- Lam, M. (004). Neural network techniques for financial performance prediction: integrating fundamental and technical analysis. *Decision Support Systems 37*(4), 567–581.
- Lameira, V. J. (003). *Negócios em Bolsas de Valores: Estratégias para Investimentos*. Alau de.
- Leonard, J. e M. A. Kramer (1990). Improvement of the backpropagation algorithm for training neural networks. *Computers chemical Engineer 14*(3), 337–341.
- Leybourne, S. J., B. P. M. McCabe, e A. R. Tremayne (1996). Can economic time series be differenced to stationarity? *Journal of Business & Economic Statistics 14*(4), 435–46.

-
- Mallat, S. (1989a). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 11(7), 674–693.
- Mallat, S. (1999). *A Wavelet Tour of Signal Processing* (ed.). Academic Press.
- Mallat, S. G. (1989b). A theory for multiresolution signal decomposition : The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7), 647–693.
- Malliavin, P. e M. E. Mancino (00). Fourier series method for measurement of multivariate volatities. *Finance and Stochastics* 6(1), 49–61.
- Matsuba, I. (1991). Neural sequential associator and its application to stock price prediction. *IECON 91 2*, 1476–1479.
- McCabe, B. P. M. e A. R. Tremayne (1995). Testing a time series for diffrence stationarity. *The Annals of Statistics* 23(3), 1015–10 8.
- Mead, R. (1990). *The Design of experiments: statistical principles for practical applications*. Cambridge University Press.
- Mertins, A. (1996). *Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications*. University of Wollongong - Australia: John Wiley Sons.
- Meyer, Y. (1993a). *Wavelets: Algorithms and Applications*. Society for Industrial and Applied Mathematic.
- Meyer, Y. (1993b). *Wavelets: Algorithms and Applications. Society for Industrial and Applied Mathematics*. Philadelphia.
- Morlet, J., A. Grossman, e R. Kronland-Martinet (1989). *Reading and Understanding Continuous Wavelet Transforms in Wavelets, Time-Frequency Methods and Phase Space*. Marseille, France: Springer.
- Murphy, J. J. (1991). *Intermarket Technical Analysis* (Wiley Finance Editions ed.). New York: John Wiley Sons.
- Nagano, M. S., E. M. Merlo, e M. C. da Silva (003). The fundamental variables and yours impacts in scrips return tax in brazil. *FAE* 6(), 13– 8.
- Niska, H., M. Heikkinen, e M. Kolehmainen (006). Genetic algorithms and sensitivity analysis applied to select inputs of a multi-layer perceptron for the prediction of air pollutant time-series. *Lecture Notes in Computer Science* 4224, 4– 31.

- O. Nerrand, P. Roussel-Ragot, L. P. G. D. (1993). Neural networks and nonlinear adaptive filtering: Unifying concepts and new algorithms. *Neural Computation* 5(), 165–199.
- Osborn, C. E. (006). *Statistical Applications for Health Information Management* (Second Edition ed.). Jones and Bartlett.
- Pereira, B. B. (1945). *Análise Espectral de Séries Temporais - Uma introdução para Engenharia, Economia e Estatística* (1 ed.). Rio de Janeiro - Brasil: Arte Final Leasing Editorial LTDA.
- Phillips, P. C. (199). Bayes models and forecasts of australian macroeconomic time series. Cowles Foundation Discussion Papers 10 4, Cowles Foundation, Yale University. available at http://ideas.repec.org/p/cwl/cwldpp/10_4.html.
- Pole, A. (1994). *Applied Bayesian Forecasting and Time Series Analysis* (1 ed.). Crc Press.
- Pole, A., M. West, e J. Harrison (1996). Applied bayesian forecasting and time series analysis. *The Statistician* 45(1), 136–137.
- Pontes, M. J. C., R. K. H. Galvão, , M. C. U. de Araújo, P. N. T. Nogueira, O. D. P. Neto, G. E. J. Teresa, e C. B. Saldanha (005). The successive projections algorithm for spectral variable selection in classification problems. *Chemometrics and Intelligent Laboratory Systems* 78(1), 11–18.
- Qi, M. (1999). Nonlinear predictability of stock returns using financial and economic variables. *Journal of Business and Economic Statistics* 17(4), 419–4 9.
- Racine, J. (001). On the nonlinear predictability of stock returns using financial and economic variables. *Journal of Business and Economic Statistics*, 19(3), 380–38 .
- Raymer, P. M. L., W. F. Goodman, L. A. Kuhn, e A. K. Jain (000). Dimensionality reduction using genetic algorithms. *IEEE Transactions on Evolutionary Computation* 4().
- Renaud, O., J. Starck, e F. Murtagh (005). Wavelet-based combined signal filtering and prediction. *IEEE Transaction on Systems, Man and Cybernetics* 35(6), 1 41–1 51.
- Riedmiller, M. e H. Braun (1993). A direct adaptive method for faster backpropagation learning: therprop algorithm. *IEEE International Conference on Neural Networks* 1, 586–591.
- Rumelhart, D. E. e J. L. McClelland (1986). *Parallel Distributed Processing*, Volume 1. The MIT Press.

-
- Russell, S. J. e P. Norvig (1995). *Artificial Intelligence: a modern approach*. New Jersey: Prentice-Hall Inc.
- Sá, G. T. (1997). *Mercado de Ações e Bolsas de Valores*. Rio de Janeiro - Brasil: Aplicação Editora Técnica Ltda.
- Saad, E. W., D. V. Prokhorov, e D. Wunsch (1998). Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. *IEEE Transactions on Neural Networks* 9(6), 1456–1470.
- Soltani, S. (00). On the use of the wavelet decomposition for time series prediction. *Neuro-computing* 48, 67– 77.
- Spall, J. (1988). *Bayesian Analysis of Time Series and Dynamic Models* (1 ed.). Marcel Dekker.
- Tavares, M. e M. Fonseca (1987). *Análise técnica - Avaliação de Investimentos*. Rio de Janeiro: IBMEC.
- Temizel, T. T. e K. Ahmad (005). Are arima neural network hybrids better than single models? *Proceedings of International Joint Conference on Neural Networks* 5, 319 –3197.
- Walczak, S., I. Yegorova, e B. H. Andrews (003). The effect of training set distributions for supervised learning artificial neural networks on classification accuracy. *Information management: support systems and multimedia technology*, 93–108.
- Wang, J.-H. e J.-Y. Leu (1996). Stock market trend prediction using arima-based neural networks. *IEEE International conference on Neural Networks* 4, 160– 165.
- Weigend, A. S., B. A. Huberman, e D. E. Rumelhart (1990). Predicting the future a connectionist approach. *International Journal of Neural Systems* 1, 193– 09.
- White, H. (1988). Economic prediction using neural networks: the case of ibm dailystock returns. *IEEE Internacional conference on Neural Networks* 2, 451–458.
- Williams, R. J. e D. Zipser (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation* 1, 70– 80.
- Yao, J., C. L. Tan, e H.-L. Poh (1999). Neural networks for technical analysis: A study on klci. *International Journal of Theoretical and Applied Finance* 2(), 1– 41.
- Zancopé, G. J. (000). *O que é bolsa de valores*. Editora Brasiliense.

Zhang, J. e A. J. Morris (1995). Dynamic process modeling using locally recurrent neural networks. *Proceedings of the American Control Conference* 4, 767– 771.

Apêndice A

Layout do arquivo de Dados da BOVESPA

Cada arquivo disponibilizado pela BOVESPA contém informações como nome e código da empresa, código da ação, código ISIN, tipo de mercado (a vista, termo, opções), especificação (ON/PN), preços (anterior, abertura, mínimo, médio, máximo, fechamento), quantidade de negócios e volume negociado com o papel, dentre outros dados disponíveis.

Após a obtenção do arquivo é necessária a utilização de um layout para a interpretação dos dados extraídos no formato .txt. É possível também a utilização de planilhas em Excel. O layout é dividido em duas partes: cabeçalho e dados.

A.0.1 Cabeçalho

Tipo de registro	Fixo “00”	N(0)	01	0
Nome do arquivo	Fixo “COTAHIST.ANO”	X(13)	03	15
Código da origem	Fixo “BOVESPA”	X(08)	16	3
Data da geração do arquivo	Formato “AAAAMMDD”	N(08)	16	3
Reserva	Preencher com brancos	X(14)	3	45

A.0.2 Dados

Nome do campo - descrição	Conteúdo	Tipo e tamanho	Pos. inicial	Pos. final
TIPREG - Tipo de registro	Fixo “01”	N(0)	01	0
Data do pregão	Formato “AAAAM-MDD”	N(0)	03	10
CODBDI - Código BDI	Utilizado para classificar as ações na emissão do boletim diário de informações	X(0)	11	1
CODNEG - Código de negociação		X(1)	13	4
TPMERC - Tipo de mercado	Cód. do mercado em que a ação está cadastrado	N(03)	5	7
NOMRES - Nome resumido da empresa		X(1)	8	39
ESPECI - Especificação do papel		X(10)	40	49
PRAZOT - Prazo em dias		X(03)	50	5
MODREF - Moeda de referência	Moeda usada na data do pregão	X(04)	53	56
PREABE - Preço de abertura da ação		(11)V99	57	69
PREMAX - Preço máximo da ação		(11)V99	70	8
PREMED - Preço médio da ação		(11)V99	83	95
PREULT - Preço do último negócio		(11)V99	96	108
PREOFC - Preço da melhor oferta de compra		(11)V99	109	1 1
PREOFV - Preço da melhor oferta de venda		(11)V99	135	147

Nome do campo - descrição	Conteúdo	Tipo e tamanho	Pos. inicial	Pos. final
TOTNFG - Número de negócios da ação no dia		N(05)	148	15
QUATOT - Quantidade de negócios total		N(18)	153	170
VOLTOT - Volume total da ação		(16)V99	171	188
PREEXE - Preço para o mercado de opções		(11)V99	189	01
INDOPC - Indicador de correção de preços		N(01)	0	0
DATVEN - Data do vencimento para mercado de opções		formato “AAAAM-MDD”	N(08)	03
FATCOT - Fator de cotação	1 para Unitária e 1000 para lote de mil	N(07)	11	17
PTOEXE - Preço para opções em dólar		(07)V06	18	30
CODISI - Código do papel no sistema ISIN		X(1)	31	4
DISMES - Número de distribuição da ação		9(03)	43	45

Apêndice B

Descrição das variáveis fundamentalista utilizadas

Valor de mercado: É o último preço pago por cada ação da empresa vezes o número total de títulos disponíveis.

Lucro líquido: É o total de vendas brutas da empresa subtraído pelo custo das mercadorias vendidas, das despesas variáveis e das despesas fixas.

Taxa de juro básica: É a taxa fixada pelo banco central, que serve de referência para o cálculo do custo do dinheiro para empréstimos bancários.

Índice BOVESPA: O Índice Bovespa é um somatório dos pesos (quantidade teórica da ação multiplicada pelo último preço da mesma) das ações integrantes de uma carteira teórica formada pelas ações mais importantes da BOVESPA. Assim sendo, pode ser apurado, a qualquer momento, por meio da Equação B.1.

$$I_{bovespa} = \sum_{i=1}^n P(i)Q(i) \quad (\text{B.1})$$

onde n é o número total de ações componentes da carteira teórica, P é o último preço da ação e Q é a quantidade teórica da ação i .

Índice de preços ao consumidor: O índice de preços ao consumidor mede a variação de preços para o consumidor com base nos gastos de quem ganha de um a vinte salários mínimos, ou seja, a grande maioria da população. Este índice é importante pois se os preços variam muito

há risco de inflação e conseqüentemente o risco de indicadores econômicos piores.

Mobilidade de capital: Mede o quanto a empresa está preparada financeiramente para atender gastos ou investimentos não programados. Uma empresa que tem mobilidade de capital alto é bem vista pelo mercado pois está preparada para eventuais emergências em casos de crises econômicas internas ou externas.

Risco país: O Risco país é um método mundialmente reconhecido que classifica o risco em se investir em um determinado país. O risco é determinado levando-se em consideração diversos indicadores econômicos, financeiros e políticos do país como por exemplo déficit fiscal, crescimento da economia, solidez das instituições, etc. O risco-país funciona como termômetro que mede a desconfiança dos investidores quanto ao cumprimento ou não do reembolso prometido pelo devedor (no caso o país) na data de vencimento dos títulos por ele emitidos. Em outras palavras, o investidor está interessado em quanto deveria ser recompensado por aplicar em papéis que embutem certa possibilidade de calote. Considerando-se o risco dos Estados Unidos como zero, a taxa paga pelo governo desse país será a base comparativa de cálculo utilizada pelo investidor. Várias instituições fazem o cálculo do risco país como por exemplo, Standard and Poors, Moody's, JP Morgan, dentre outros.

Apêndice C

Software de Predição

O software de predição foi criado a partir da necessidade de testar as técnicas propostas no problema de predição de série temporais econômicas. Em uma primeira versão, o software era executado em linha de comando, sendo que a interface gráfica apresentada na figura C.1, foi desenvolvida especialmente para a apresentação nesta dissertação.

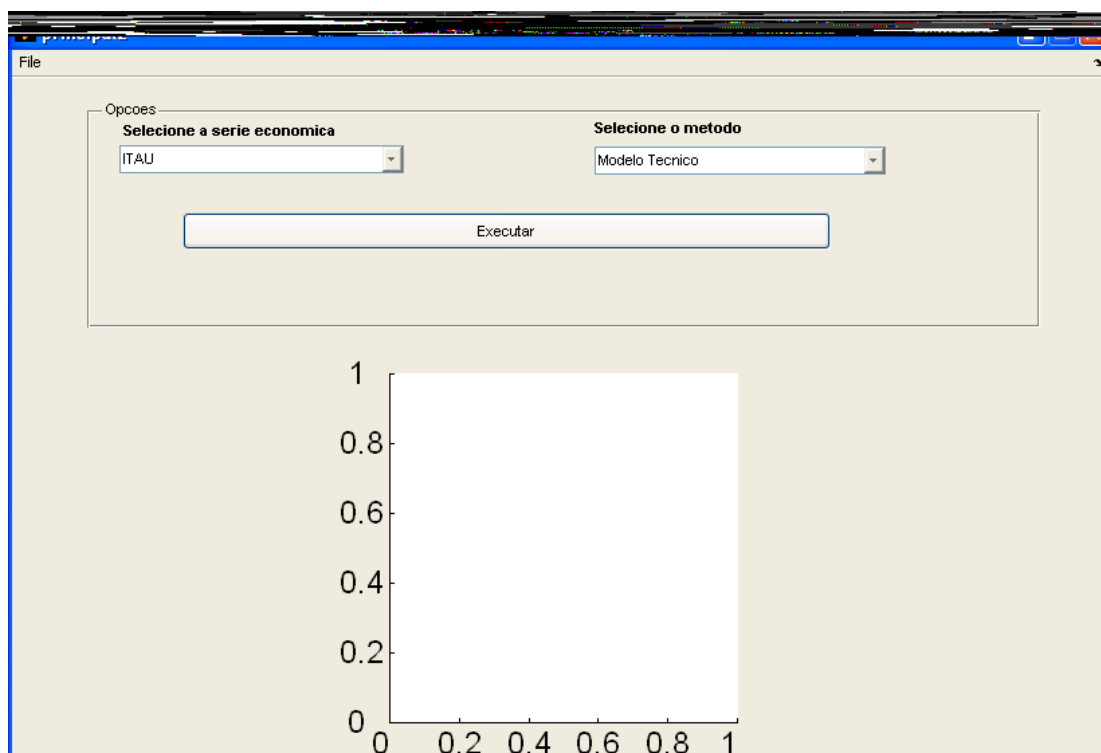


Figura C.1: Software de predição desenvolvido no MatLab.

O menu de seleção da série econômica, permite ao usuário selecionar qual série será utilizada

no software. O menu de método, permite ao usuário selecionar qual tipo de método será empregado na predição, sendo que o resultado será plotado em um gráfico de resultado esperado *versus* resultado do método.

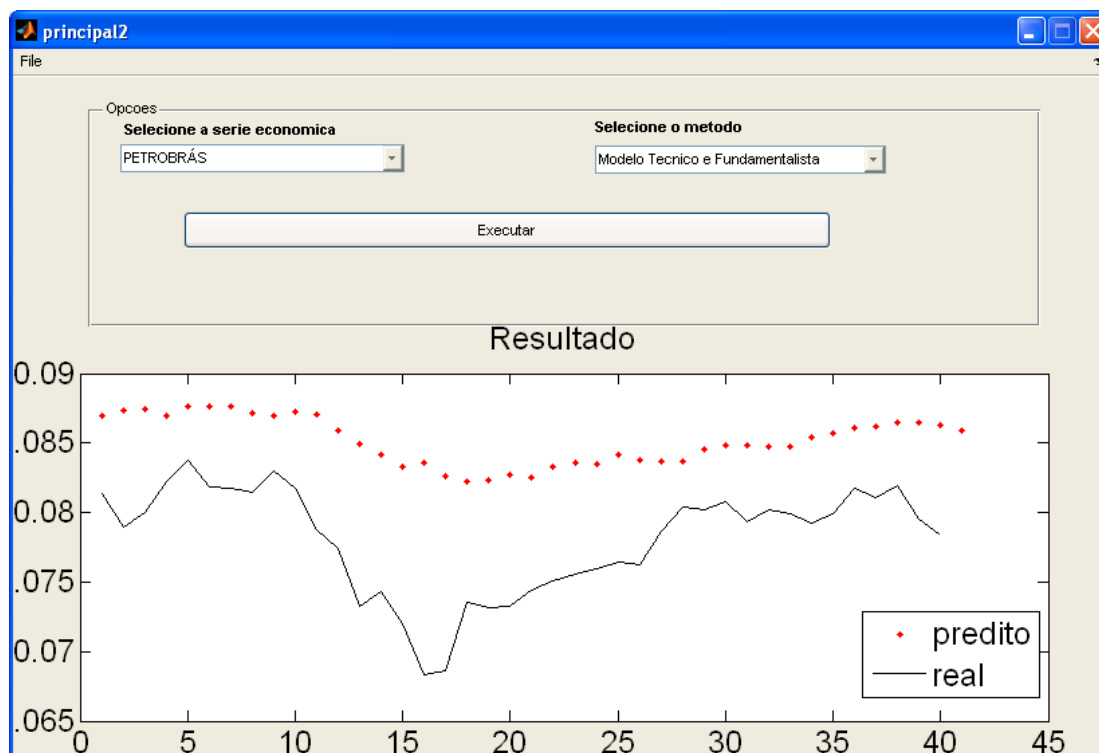


Figura C. : Exemplo de Resultado Obtido no Software de predição.