



Luiza Maria Oliveira da Silva

**Uma Aplicação de Árvores de Decisão, Redes
Neurais e KNN para a Identificação de Modelos
ARMA Não-Sazonais e Sazonais**

Tese de Doutorado

Tese apresentada ao Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio como parte dos requisitos parciais para obtenção do título de Doutor em Engenharia Elétrica da PUC-Rio.

Orientador: Prof. Reinaldo Castro Souza

Rio de Janeiro
Setembro de 2005



Luiza Maria Oliveira da Silva

**Uma Aplicação de Árvores de Decisão, Redes
Neurais e KNN para a Identificação de Modelos
ARMA Não-Sazonais e Sazonais**

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Dr. Reinaldo Castro Souza

Orientador

Departamento de Engenharia Elétrica - PUC - Rio

Dr. Flávio Joaquim de Souza

UERJ

Dra. Maria Augusta Soares Machado

IBMEC- RJ

Dr. Ricardo Tanscheit

Departamento de Engenharia Elétrica – PUC-Rio

Dr. Antônio de Araujo Freitas Junior

FGV - RJ

Prof. José Eugenio Leal

Coordenador(a) Setorial do Centro

Técnico Científico - PUC-Rio

Rio de Janeiro, 5 de setembro de 2005

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Luiza Maria Oliveira da Silva

Graduada em Matemática pela Universidade Federal Fluminense em 1987, mestre em Engenharia de produção pela Universidade Federal Fluminense em 1996 e doutora pela Pontifícia Universidade Católica do Rio de Janeiro em Engenharia Elétrica na área de Métodos de Apoio à Decisão em 2005. Professora das Faculdades IBMEC-RJ desde 1995 ministrando aulas de Matemática no curso de graduação. Tem livros e artigos publicados nas áreas de Matemática, Estatística, Pesquisa Operacional e Inteligência Computacional Aplicada.

Ficha catalográfica

Silva, Luiza Maria Oliveira da

Uma aplicação de árvores de decisão, redes neurais e KNN para a identificação de modelos ARMA não sazonais e sazonais / Luiza Maria Oliveira da Silva ; orientador: Reinaldo Castro Souza. – Rio de Janeiro : PUC-Rio, Departamento de Engenharia Elétrica, 2005.

145 f. ; 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica.

Inclui referências bibliográficas

1. Engenharia Elétrica – Teses. 2. Séries temporais. 3. Identificação dos modelos Box & Jenkins não sazonais e sazonais. 4. Árvores de decisão. 5. Redes neurais. 6. K-Nearest Neighbor (KNN). I. Souza, Reinaldo Castro. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica . III. Título.

CDD: 621.3

Agradecimentos

À Deus por ter me concedido sabedoria e perseverança na elaboração desta tese.

Aos meus pais e irmã que sempre me apoiaram e torceram por mim.

Ao meu marido Sérgio Hosken, que nos momentos mais difíceis, sempre me incentivava para que eu continuasse lutando e nunca me esquecesse do meu objetivo.

Ao meu professor e orientador Reinaldo Castro Souza pela oportunidade de realizar este trabalho, pela sua orientação e paciência.

Ao meu co-orientador, o professor Flávio Joaquim de Souza, que com sua paciência e amizade muito me orientou, pois a sua ajuda foi imprescindível na construção deste trabalho.

À professora Maria Augusta Soares Machado pelo incentivo, apoio, orientação e, sobretudo, a amizade.

Ao professor Ricardo Tanscheit por sua valiosa ajuda.

À Alcina, Márcia e Ana pelo auxílio nos processos burocráticos.

À Ana Paiva e Flávio Nascimento que me receberam com muito carinho.

À amiga Christiane Romeo pela atenção especial na revisão deste trabalho e pelo apoio em todas as horas.

À todas as pessoas que, durante este período de minha vida, me ajudaram direta ou indiretamente.

Resumo

Silva, Luiza Maria O.; Souza, Reinaldo Castro (Orientador). **Uma Aplicação de Árvores de Decisão, Redes Neurais e KNN para a Identificação de Modelos ARMA Não-Sazonais e Sazonais**. Rio de Janeiro, 2005. 145p. Tese de Doutorado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

A metodologia Box & Jenkins tem sido mais utilizada para fazer previsões do que outros métodos até então. Alguns analistas têm relutado, entretanto, em usar esta metodologia, em parte porque a identificação da estrutura adequada é uma tarefa complexa. O reconhecimento tanto dos padrões de comportamento das funções de autocorrelação quanto da autocorrelação parcial (teórica/estimada) dependem da série temporal através da qual é possível extraí-las. Uma vez obtidos os resultados, pode-se inferir qual o tipo de estrutura Box & Jenkins adequada para a série. A proposta do trabalho é desenvolver três novas metodologias de identificação automática das estruturas Box & Jenkins ARMA simples e/ou sazonais, identificar os filtros sazonal e linear da série de uma forma menos complexa. A primeira metodologia utiliza árvores de decisão, a segunda, redes neurais e a terceira, K-Nearest Neighbor (KNN). A estas metodologias serão utilizadas as estruturas Box & Jenkins sazonais de períodos 3, 4, 6 e 12 e não sazonais. Os resultados são aplicados a séries simuladas, bem como a séries reais. Como comparação, utilizou-se o método automático de identificação proposto no software FPW-XE.

Palavras-chave

Séries temporais; identificação dos modelos Box & Jenkins não sazonais e sazonais; árvores de decisão; redes neurais; K-Nearest Neighbor (KNN).

Abstract

Silva, Luiza Maria O.; Souza, Reinaldo Castro (Advisor). **The Use of Decision Trees, Neural Networks and K - Nearest Neighbors Systems to Automatically Identify Box & Jenkins Non-Seasonal and Seasonal Structures.** Rio de Janeiro, 2005. 145p. Tese de Doutorado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The Box & Jenkins is the most popular forecasting technique. However, some researchers have not embraced it because the identification of its structure is highly complex. The process of proper characterizing the properties of both autocorrelation functions and partial correlation (theoretical or estimated) depends on the time series from which they are being obtained. Given the results in question, it is possible to infer the proper Box & Jenkins structure for the time series being studied. For the reasons above, the goal of this dissertation is to develop three new methodologies to identifying, in an automatic fashion, the Box & Jenkins structure of an ARMA series. The methodologies identify, in a simpler manner, both the seasonal and linear filters of the series. The first methodology applies the decision tree. The second applies the neural networks. The third applies the K-Nearest Neighbor (KNN). In each of them the Box & Jenkins seasonal structures of 3, 4, 6 and 12 periods were used, as well as the non-seasonal structure. The results are applied to simulated and actual series. For comparison purposes, the automatic identification procedure of the software FPW-XE is also used.

Keywords

Time series; identifying the Box & Jenkins structure of an ARMA series non-seasonal and seasonal; decision tree; neural network; K-Nearest Neighbor (KNN).

Sumário

1 INTRODUÇÃO	13
1.1. Motivação	13
1.2. Objetivo	14
1.3. Organização do trabalho	14
2 METODOLOGIA DE PREVISÃO DE SÉRIES TEMPORAIS - BOX & JENKINS	15
2.1. Introdução	15
2.2. Possíveis Modelos Não sazonais na Metodologia de Box & Jenkins	16
2.2.1. Modelo Auto-regressivo (AR)	16
2.2.2. Modelo Médias Móveis (MA)	17
2.2.3. Modelo Auto-Regressivo Médias Móveis (ARMA)	17
2.2.4. Modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA)	18
2.2.5. Etapas da Metodologia de Box & Jenkins	18
2.3. Sazonalidade	32
2.4. Possíveis Modelos Puramente Sazonais na Metodologia de Box & Jenkins	32
2.4.1. Modelo Sazonal Auto-Regressivo de Ordem P – SAR(P) _s (<i>Seasonal Autoregressive</i>)	32
2.4.2. Modelo Sazonal de Médias Móveis de Ordem Q – SMA(Q) _s (<i>Seasonal Moving Average</i>)	33
2.4.3. Modelo Sazonal Auto-Regressivo de Médias Móveis – SARMA(P,Q) _s (<i>Seasonal Autoregressive Moving Average</i>)	33
2.4.4. Modelo Sazonal Auto-Regressivo Integrado de Médias Móveis – SARIMA(P,D,Q) _s (<i>Seasonal Autoregressive Integrated Moving Average</i>)	33
2.5. Possíveis Modelos Sazonais na Metodologia de Box & Jenkins	34
2.5.1. Modelo Sazonal Auto-Regressivo Integrado de Médias Móveis – SARIMA (p,d,q)x (P,D,Q) _s	34
2.5.2. Exemplo de Modelo SARIMA - Modelo <i>Airline</i>	37
2.6. Componentes Estruturais de um Modelo	37
2.6.1. Sazonalidade Estocástica	38
2.6.2. Sazonalidade Determinística	39
2.7. Testes estatísticos para verificação da validade do modelo	40
3 ÁRVORE DE DECISÃO	41
3.1. Introdução	41
3.2. Representação de uma árvore de decisão	41

3.3. Entropia	42
3.4. Índice Gini	44
3.5. O problema do <i>overfitting</i>	44
3.6. Podagem	46
3.7. Algoritmos	47
4 REDES NEURAIIS	49
4.1. Introdução	49
4.2. Arquitetura da rede	49
4.3. Processos de Aprendizagem	50
4.4. Treinamento de uma rede neural	51
4.5. O Algoritmo Retropropagação (<i>backpropagation</i>)	51
5 KNN (K – <i>Nearest Neighbors</i>)	53
5.1. Introdução	53
5.2. Metodologia	53
6 APLICAÇÃO DA METODOLOGIA PROPOSTA	56
6.1. Introdução	56
6.2. Metodologia aplicada à árvore de decisão	58
6.3. Metodologia aplicada às redes neurais	58
6.4. Metodologia aplicada a KNN (<i>K-Nearest Neighbors</i>)	60
6.5. Resultados obtidos	60
6.5.1. Resultados obtidos ao utilizar Árvore de Decisão	61
6.5.1.1. Análise dos resultados para os modelos Box & Jenkins não-sazonais.	61
6.5.1.2. Análise dos resultados para os modelos Box & Jenkins sazonais.	62
6.5.2. Resultados obtidos ao utilizar Redes Neurais	69
6.5.2.1. Análise dos resultados para os modelos Box & Jenkins não-sazonais.	69
6.5.2.2. Análise dos resultados para os modelos Box & Jenkins sazonais.	70
6.5.3. Resultados obtidos ao utilizar KNN – distância Euclidiana	74
6.5.3.1. Análise dos resultados para os modelos Box & Jenkins não-sazonais.	74
6.5.3.2. Análise dos resultados para os modelos Box & Jenkins sazonais.	75
6.5.4. Resultados obtidos ao utilizar KNN – distância Manhattan	77
6.5.4.1. Análise dos resultados para os modelos Box & Jenkins não-sazonais.	77
6.6. Conclusões	80
6.7. Validação do classificador KNN	86
6.8. Validação do KNN utilizando a série <i>AIRLINE</i>	87
7 CONCLUSÕES	90
7.1. Trabalhos futuros	90
8 REFERÊNCIAS BIBLIOGRÁFICAS	91

9 ANEXOS	97
ANEXO 1 - RESULTADOS COMPLEMENTARES OBTIDOS AO UTILIZAR “ÁRVORES DE DECISÃO”	97
ANEXO 2 - RESULTADOS COMPLEMENTARES OBTIDOS AO UTILIZAR REDES NEURAIAS	117
ANEXO 3 - RESULTADOS COMPLEMENTARES OBTIDOS AO UTILIZAR KNN – DISTÂNCIA EUCLIDIANA	131
ANEXO 4 - RESULTADOS COMPLEMENTARES OBTIDOS AO UTILIZAR KNN – DISTÂNCIA MANHATTAN	138

Lista de figuras

Figura 1 - ACF e PACF do modelo AR(1)	25
Figura 2 - ACF e PACF do modelo AR(2)	26
Figura 3 - Região de admissibilidade em função de ϕ_1 e ϕ_2 para o modelo AR(2)	27
Figura 4 - Região de admissibilidade em função de ρ_1 e ρ_2 para o modelo AR(2)	27
Figura 5 - ACF e PACF do modelo MA(1)	28
Figura 6 - ACF e PACF do modelo MA(2)	28
Figura 7 - Região de admissibilidade em função de θ_1 e θ_2 para o modelo MA(2)	29
Figura 8 - Região de admissibilidade em função de ρ_1 e ρ_2 para o modelo MA(2)	29
Figura 9 - ACF e PACF do modelo ARMA(1,1)	30
Figura 10 - Região de admissibilidade em função de θ_1 e ϕ_1 para o modelo ARMA(1, 1)	31
Figura 11 - Região de admissibilidade em função de ρ_1 e ρ_2 para o modelo ARMA(1, 1)	31
Figura 12 - Representação de uma árvore de decisão e sua respectiva representação no espaço	42
Figura 13 - Representação gráfica $p(x)$ versus entropia	43
Figura 14 - Representação gráfica o percentual de erro no conjunto de treinamento e teste versus o número de nós da árvore	45
Figura 15 - Exemplo de podagem	46
Figura 16 - Representação de uma rede neural com m nós na camada de entrada, 2 camadas ocultas de neurônios e uma camada de saída	50
Figura 17 - Sinais funcionais e de erro numa rede neural	52
Figura 18 - Classificação pelo método KNN	54
Figura 19 - Classificação pelo método KNN utilizando hiper-esferas de raio R	55
Figura 20 - Representação da arquitetura da rede neural	59
Figura 21 - Árvore de decisão para modelos não-sazonais	61
Figura 22 - Árvore de decisão para modelos sazonais de período 3	63
Figura 23 - Árvore de decisão para modelos sazonais de período 4	64
Figura 24 - Árvore de decisão para modelos sazonais de período 6	66
Figura 25 - Árvore de decisão para modelos sazonais de período 12	68
Figura 26 - Gráfico da série AIRLINE antes da diferenciação	88
Figura 27 - Gráfico da série AIRLINE após a diferenciação	88
Figura 28 - Árvore de decisão para modelos não sazonais sem ACF5 e PACF5	97
Figura 29 - Árvore de decisão para modelos não sazonais (conjunto modificado)	98
Figura 30 - Árvore de decisão para modelos não sazonais sem	

ACF5 e PACF5 (conjunto modificado)	100
Figura 31 - Árvore de decisão para modelos sazonais de período 3 (conjunto modificado)	101
Figura 32 - Árvore de decisão para modelos sazonais de período 4 (conjunto modificado)	103
Figura 33 - Árvore de decisão para modelos sazonais de período 6 (conjunto modificado)	104
Figura 34 - Árvore de decisão para modelos sazonais de período 12 sem ACF60 e PACF60	106
Figura 35 - Árvore de decisão para modelos sazonais de período 12 (conjunto modificado)	107
Figura 36 - Árvore de decisão para modelos sazonais de período 12 sem ACF60 e PACF60 (conjunto modificado)	109
Figura 37 - Árvore de decisão para modelos sazonais de período 12 onde $ \Theta > 0,2$ e $ \Phi > 0,2$	110
Figura 38 - Árvore de decisão para modelos sazonais de período 12 onde $-0,2 < \Theta < 0,2$ e $-0,2 < \Phi < 0,2$	112
Figura 39 - Árvore de decisão para modelos sazonais de período 12 onde $a_t \sim N(0,2)$	113
Figura 40 - Árvore de decisão para modelos sazonais de período 12 onde $a_t \sim N(0,2)$ (conjunto modificado)	115

Lista de tabelas

Tabela 1 - Comportamento teórico dos modelos AR(p), MA(q), ARMA(p,q), SAR(P), SMA(Q) e SARMA(P,Q)	36
Tabela 2 - Resultados obtidos utilizando o classificador árvore de decisão	80
Tabela 3 - Resultados obtidos utilizando o classificador redes neurais	81
Tabela 4 - Resultados obtidos utilizando o classificador KNN – distância Euclidiana	82
Tabela 5 - Resultados obtidos utilizando o software FPW	83
Tabela 6 - Porcentagem média de acerto em cada metodologia	83
Tabela 7 - Percentual de AR(1), MA(1), SAR(1) e SMA(1) encontrados no ARMA(1,1) e SARMA(1,1) utilizando o KNN	84
Tabela 8 - Percentual de AR(1), MA(1), SAR(1) e SMA(1) encontrados no ARMA(1,1) e SARMA(1,1) utilizando o FPW	85
Tabela 9 - Total percentual de acertos encontrados no ARMA(1,1) e SARMA(1,1) utilizando o KNN	85
Tabela 10 - Total percentual de acertos encontrados no ARMA(1,1) e SARMA(1,1) utilizando o FPW	86
Tabela 11 - Classificação de alguns modelos Box & Jenkins	87