



Bernardo da Rocha Spindel

**Análise e Desenvolvimento de Sistema de
Estimação de Modelos da Classe STAR-TREE**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio.

Orientador: Prof. Álvaro de Lima Veiga Filho

Rio de Janeiro
Agosto de 2008



Bernardo da Rocha Spindel

**Análise e Desenvolvimento de Sistema de
Estimação de Modelos da Classe STAR-TREE**

Dissertação de Mestrado apresentada como requisito parcial para obtenção do grau de Mestre 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. Álvaro de Lima Veiga Filho
Orientador

Departamento de Engenharia Elétrica – PUC-Rio

Dr. Joel Maurício Corrêa da Rosa
UFPR

Dr. Cristiano Augusto Coelho Fernandes
Departamento de Economia – PUC-Rio

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico

Rio de Janeiro, 29 de agosto de 2008

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

Bernardo da Rocha Spindel

Bernardo Spindel é formado em Engenharia de Produção e mestrando em Métodos de Apoio à decisão com ênfase em estatística pela PUC-Rio. Suas áreas de interesse incluem a análise de séries temporais, métodos de regressão e estatística multivariada.

Ficha Catalográfica

Spindel, Bernardo da Rocha

Análise e desenvolvimento de sistema de estimação de modelos da classe STAR-TREE / Bernardo da Rocha Spindel ; orientador: Álvaro de Lima Veiga Filho. – 2008.

89 f. ; 30 cm

Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. STAR-TREE. 3. Otimização. 4. Algoritmos genéticos. 5. AR. 6. Função logística. 7. Pertinência. I. Veiga Filho, Álvaro de Lima. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

Agradecimentos

À minha família, pela educação, carinho e apoio, sem os quais este trabalho não poderia ter sido concluído.

Aos amigos Bruno Flach, Alexandre Street e Eduardo Faria por toda a colaboração, orientação e discussões que ao longo de todo o mestrado, me ajudaram a aprofundar meus conhecimentos.

Aos amigos Alexandre Donato, Sérgio Monteiro e Felipe Miana, que de diferentes formas e em diferentes áreas, contribuíram bastante para a realização desta dissertação.

Ao amigo Alexander Virtser, pelo incentivo e oportunidade de aprender bastante ao longo do tempo de uma pessoa de profundo conhecimento teórico e prático.

À Andréa Ribeiro Soares, por toda ajuda e incentivo dados ao longo de todo esse processo.

Ao orientador Álvaro Veiga, pelo estímulo e apoio.

À CNPq e à PUC-Rio, pelos auxílios concedidos.

Resumo

Spindel, Bernardo da Rocha; Veiga Filho, Álvaro de Lima (Orientador). **Análise e Desenvolvimento de Sistema de Estimação de Modelos da Classe STAR-TREE**. Rio de Janeiro, 2008, 89p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Na análise de séries temporais, os modelos lineares amplamente difundidos e utilizados, como regressões lineares e modelos auto-regressivos, não são capazes de capturar sua natureza muitas vezes não-linear, oferecendo resultados insatisfatórios. Séries financeiras, por exemplo, apresentam este tipo de comportamento. Ao longo dos últimos anos, houve o surgimento de muitos modelos não lineares para análise de séries temporais, tanto estatísticos como de inteligência computacional, baseados em redes neurais. Esta dissertação se propõe a analisar a performance do modelo STAR-Tree sob diversos cenários de configuração, parametrização e metodologias de estimação. Esta classe de modelos subdivide os dados de uma série temporal em regiões distintas que atendem critérios especificados em funções chamadas de pertinências. A cada região é atribuído um modelo linear auto-regressivo. Cada dado estimado pode estar em alguma das regiões com algum grau de pertinência determinado pelas funções fornecidas pelo modelo principal. Fatores como a proximidade das regiões, a suavidade das funções de pertinência e a falta de diversidade nos dados podem dificultar a estimação dos modelos. Para avaliar a qualidade das estimações sob os diversos cenários, foi construído um sistema capaz de gerar séries artificiais, importar séries externas, estimá-las sob a modelagem STAR-Tree, e gerar simulações de Monte Carlo que avaliam a qualidade da estimação de parâmetros e a capacidade de detecção das estruturas de árvore do modelo. Ele foi utilizado como ferramenta para realizar as análises presentes na dissertação, e permitiu que se testassem diferentes configurações de métodos e parametrizações com facilidade.

Palavras-chave

Engenharia Elétrica, Smooth Transition Auto Regressive Tree, STAR-Tree, Otimização, Algoritmos Genéticos, AR, Função de Logística, Pertinência.

Abstract

Spindel, Bernardo da Rocha; Veiga Filho, Álvaro de Lima (Advisor). **Analysis and Development of a STAR-Tree Model Estimation Software**. Rio de Janeiro, 2008, 89p. MSc Dissertation – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

In time series analysis, linear models that have been broadly used, such as linear regressions and auto-regressive models, are not able to capture the some times non linear nature of some data, offering poor estimation results. Financial series, for instance, show that kind of behavior. Over the last years, a great number of non linear models have been developed in order to analyze time series, some of them statistical, others based on computational intelligence techniques such as neural networks. The purpose of this dissertation is to analyze the performance of the STAR-Tree model under distinct scenarios that differ in model specification, parameterization and estimation methodologies. This class of models splits time series data into individual regions which fulfill the criteria set up by functions called pertinences. A linear model then is selected for each one of those regions. Each estimated data point can belong to one of the mentioned regions with some degree of pertinence, supplied by the above mentioned pertinence functions. Aspects like the proximity between regions, the smoothness of the pertinence functions and the lack of diversity in real data can significantly affect the estimation of models. In order to evaluate the quality of the estimations under the different proposed scenarios, a software was developed with the capabilities of generating artificial time series, importing external series, estimating them under the STAR-Tree model, and generating Monte Carlo simulations that evaluate the quality of parameter estimation and the tree structure detection capability of the model. The software was used as the single tool to generate this dissertation's analyses, and allowed that different model specifications and methods could be tested without difficulty.

Keywords

Electrical Engineering, Smooth Transition Auto Regressive Tree, STAR-Tree, Optimization, Genetic Algorithms, AR, Logistic Function, Pertinence

Sumário

1	INTRODUÇÃO	12
2	MODELOS DA CLASSE STAR ESTRUTURADOS EM ÁRVORE	14
2.1.	Introdução	14
2.2.	Formulação Geral dos Modelos STAR	14
2.3.	O Modelo LSTAR	16
2.4.	Árvores de Decisão	18
2.5.	O Modelo STAR-Tree	22
3	ESTIMAÇÃO DE PARÂMETROS DO MRSTAR	24
3.1.	Introdução	24
3.2.	Dificuldades na estimação	26
3.3.	Método de Gradiente	28
3.4.	Método de Newton	30
3.5.	BFGS	32
3.6.	Aplicação de Métodos de Otimização na Estimação do MRSTAR	36
3.7.	Estimação por Algoritmos Genéticos	37
3.8.	Resultados	40
3.8.1.	Média e desvio padrão do MSE	41
3.8.2.	Pertinência média dos nós terminais	42
3.8.3.	Estimação dos parâmetros não-lineares	46
3.8.4.	Estimação dos parâmetros lineares	50
4	APLICAÇÃO DO MODELO STAR-TREE	57
4.1.	Introdução	57
4.2.	Método CART	58
4.3.	Estimação por teste de hipótese	62
4.4.	Aplicação a dados reais	68
4.4.1.	Modelo ARX	70
4.4.2.	Modelo STAR-Tree	72
4.5.	Conclusões	78
5	DESCRIÇÃO DO SISTEMA	80
5.1.	Objetivo	80
5.2.	Módulo de estimação de séries temporais	80
5.2.1.	Parâmetros de estimação	81
5.2.2.	Gerador de séries artificiais	83
5.2.3.	Estimação e visualização de resultados	85
5.2.4.	Simulação através de modelos gerados artificialmente	88
6	REFERÊNCIAS BIBLIOGRÁFICAS	89

Lista de Figuras

Figura 2.1	Função Logística	15
Figura 2.2	Função Exponencial	15
Figura 2.3	Função Degrau	15
Figura 2.4	Gama fixo com valor baixo, c variando	17
Figura 2.5	Gama fixo com valor alto, c variando	17
Figura 2.6	Exemplo de regiões de dados distintas com modelos específicos de estimação	18
Figura 2.7	Exemplo de árvore de decisão	19
Figura 2.8	Ciclo de estimação de uma árvore de decisão	21
Figura 3.1	LSTAR(1) com $\gamma=0$ e $c=0$	26
Figura 3.2	LSTAR(1) com $\gamma=10$ e $c=0$	26
Figura 3.3	Superfície da função $l(Y \Phi)$ para diferentes valores de c e γ . O modelo gerado foi um LSTAR(1) com $\gamma = 0, c = 0, B_1 = [0 \quad -.8], B_2 = [0 \quad .6]$, gerado com 1000 pontos	27
Figura 3.4	Superfície da função $l(Y \Phi)$ para diferentes valores de c e γ . O modelo gerado foi um LSTAR(1) com $\gamma = 50, c = 0, B_1 = [0 \quad -.8], B_2 = [0 \quad .6]$, gerado com 1000 pontos	28
Figura 3.5	Ilustração do método do Gradiente	30
Figura 3.6	Ilustração comparando o método de Newton com o do Gradiente	32
Figura 3.7	Ilustração do método da secante convergindo para o mínimo da função, partindo de dois pontos inicialmente selecionados	34
Figura 3.8	Diferentes métodos estimando o próximo passo da matriz B	35
Figura 3.9	Algoritmo de estimação de parâmetros por métodos de otimização	37
Figura 3.10	Histogramas de pertinências estimadas para um dado nó sob o método BFGS para $\gamma=1, \gamma=5$ e $\gamma=10$	44
Figura 3.11	Histogramas de pertinências estimadas para um dado nó sob algoritmo genético para $\gamma=1, \gamma=5$ e $\gamma=10$	45
Figura 3.12	Dispersão dos parâmetros de localização e suavidade para $\beta_0 = \begin{bmatrix} 0 \\ 0.2 \end{bmatrix}; \beta_1 = \begin{bmatrix} 0 \\ 0.9 \end{bmatrix}$, estimado através do método BFGS e	46

- $\gamma=5$
- Figura 3.13 Dispersão dos parâmetros de localização e suavidade para $\beta_0 = \begin{bmatrix} 0 \\ -0.5 \end{bmatrix}; \beta_1 = \begin{bmatrix} 0 \\ 0.7 \end{bmatrix}$, estimado através do método BFGS e $\gamma=5$ 46
- Figura 3.14 Dispersão dos parâmetros de localização e suavidade para $\beta_0 = \begin{bmatrix} 0 \\ 0.5 \\ -0.2 \end{bmatrix}; \beta_1 = \begin{bmatrix} 0 \\ -0.4 \\ 0.3 \end{bmatrix}$, estimado através do método BFGS e $\gamma=5$ 47
- Figura 3.15 Dispersão dos parâmetros de localização e suavidade para $\beta_0 = \begin{bmatrix} 0 \\ 0.2 \end{bmatrix}; \beta_1 = \begin{bmatrix} 0 \\ 0.9 \end{bmatrix}$, estimado através do método de Newton e $\gamma=5$ 47
- Figura 3.16 Dispersão dos parâmetros de localização e suavidade para $\beta_0 = \begin{bmatrix} 0 \\ -0.5 \end{bmatrix}; \beta_1 = \begin{bmatrix} 0 \\ 0.7 \end{bmatrix}$, estimado através do método de Newton e $\gamma=5$ 47
- Figura 3.17 Dispersão dos parâmetros de localização e suavidade para $\beta_0 = \begin{bmatrix} 0 \\ 0.2 \end{bmatrix}; \beta_1 = \begin{bmatrix} 0 \\ 0.9 \end{bmatrix}$, estimado através do método do Gradiente e $\gamma=5$ 48
- Figura 3.18 Dispersão dos parâmetros de localização e suavidade para $\beta_0 = \begin{bmatrix} 0 \\ -0.5 \end{bmatrix}; \beta_1 = \begin{bmatrix} 0 \\ 0.7 \end{bmatrix}$, estimado através do método do Gradiente e $\gamma=5$ 48
- Figura 3.19 Dispersão dos parâmetros de localização e suavidade para $\beta_0 = \begin{bmatrix} 0 \\ 0.2 \end{bmatrix}; \beta_1 = \begin{bmatrix} 0 \\ 0.9 \end{bmatrix}$, estimado através de algoritmo genético e $\gamma=5$ 48
- Figura 3.20 Dispersão dos parâmetros de localização e suavidade para $\beta_0 = \begin{bmatrix} 0 \\ -0.5 \end{bmatrix}; \beta_1 = \begin{bmatrix} 0 \\ 0.7 \end{bmatrix}$, estimado através de algoritmo genético e $\gamma=5$ 49

Figura 3.21	Dispersão dos parâmetros de localização e suavidade para $\beta_0 = \begin{bmatrix} 0 \\ -0.5 \end{bmatrix}; \beta_1 = \begin{bmatrix} 0 \\ 0.4 \end{bmatrix}; \beta_2 = \begin{bmatrix} 0 \\ 0.7 \end{bmatrix}$, estimado através de algoritmo genético e $\gamma=5$	51
Figura 3.22	Estimação dos parâmetros lineares através do método BFGS para o modelo com $\beta_0 = \begin{bmatrix} 0 \\ -0.5 \end{bmatrix}; \beta_1 = \begin{bmatrix} 0 \\ 0.7 \end{bmatrix}$ e $\gamma=5$	52
Figura 3.23	Valores detectados para os parâmetros lineares para cada uma das pertinências encontradas através do método BFGS para o modelo com $\beta_0 = \begin{bmatrix} 0 \\ -0.5 \end{bmatrix}; \beta_1 = \begin{bmatrix} 0 \\ 0.7 \end{bmatrix}$ e $\gamma=5$	54
Figura 4.1	Ciclo de detecção de estrutura da árvore STAR	57
Figura 4.2	Exemplo de CART de classificação	59
Figura 4.3	Série de consumo médio horário de combustível, de janeiro a fevereiro de 2005	69
Figura 4.4	Variáveis selecionadas X série de consumo	70
Figura 4.5	Série de resíduos da estimação	71
Figura 4.6	Histograma dos resíduos da estimação	71
Figura 4.7	Probabilidade empírica acumulada dos resíduos	72
Figura 4.8	Árvore estimada pelo modelo STAR-Tree	73
Figura 4.9	Pertinência acumulada capturada pelos nós de decisão a cada dado da série	74
Figura 4.10	Funções de pertinência resultantes da estimação	75
Figura 4.11	Série de resíduos da estimação do modelo STAR	76
Figura 4.12	Histograma dos resíduos da estimação	76
Figura 4.13	Probabilidade empírica acumulada dos resíduos	77
Figura 4.14	Estimação do modelo STAR-Tree	78
Figura 5.1	Escolha do método de estimação dos parâmetros não lineares	81
Figura 5.2	Exemplo de parametrização para algoritmos genéticos	82
Figura 5.3	Escolha do método de crescimento da árvore	83
Figura 5.4	Escolha dos parâmetros de um nó de decisão	84
Figura 5.5	Escolha dos parâmetros de um nó de terminal	84
Figura 5.6	Geração de séries a partir de modelos previamente montados	95
Figura 5.7	Processo de estimação da série	86
Figura 5.8	Exemplo de estimação de série	86
Figura 5.9	Funções de pertinência nos modelos gerador e estimado	87
Figura 5.10	Análise dos resíduos decorrentes da estimação	87
Figura 5.11	Exemplo de relatório gerado pelo módulo de simulações	88

Lista de Tabelas

Tabela 4.1	Descrição das estatísticas de avaliação dos modelos	69
Tabela 4.2	Parâmetros resultantes da estimação	70
Tabela 4.3	Estatísticas de avaliação do modelo ARX	72
Tabela 4.4	Parâmetros lineares resultantes da estimação do modelo STAR- Tree	75
Tabela 4.5	Estatísticas de avaliação do modelo STAR-Tree	77