

UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA PREVISÃO DE SÉRIES TEMPORAIS

Aida Araújo Ferreira(1); Elvis Gonçalves de Lira(2)

(1) Instituto Federal de Educação, Ciência e Tecnologia (IFPE), Av. Prof Luiz Freire, 500 Cidade Universitária, CEP: 50740-540 – Recife – PE – Brasil, aida_mary2@yahoo.com.br

(2) Instituto Federal de Educação, Ciência e Tecnologia (IFPE), Av. Prof Luiz Freire, 500 Cidade Universitária, CEP: 50740-540 – Recife – PE – Brasil, elvislira@yahoo.com.br

RESUMO

Diante da importância da previsão da velocidade dos ventos para a previsão da geração da energia eólica e também diante da complexidade e dinamismo dessa tarefa, o presente trabalho apresenta os resultados da criação de um sistema inteligente híbrido para previsão de séries temporais, através de um estudo de caso da previsão em curto prazo da velocidade média horária dos ventos para a cidade de Belo Jardim no estado de Pernambuco. Este trabalho apresenta também os resultados obtidos com a utilização dos métodos estatísticos de amortização exponencial e média móvel, para previsão da mesma série temporal. No método de amortização exponencial toda a série é usada para a previsão dos dados enquanto que na média móvel faz-se uso apenas de uma parte dos dados. O objetivo deste trabalho é mostrar as vantagens do uso de redes neurais artificiais na previsão de séries temporais quando comparado à utilização de métodos estatísticos tradicionais.

Palavras-chave: Previsão de séries temporais, redes neurais artificiais, amortização exponencial, média móvel.

1 INTRODUÇÃO

Em todas as áreas do conhecimento humano nos deparamos com séries temporais, por exemplo:

- A cotação diária do dólar no período de um mês;
- Consumo mensal de determinado alimento em cinco anos;
- Quantidades de chuvas (mensalmente), para uma determinada região, em 50 anos;

Barros (2007) define série temporal como sendo um conjunto de observações ordenadas no tempo (nem sempre igualmente espaçadas), e que apresentam dependência serial, isto é, dependência entre instantes de tempo.

Vários são os métodos usados para previsão de séries temporais, porém o uso de redes neurais vem sendo bastante difundido.

Uma rede neural artificial (RNA) é composta por unidades de processamento (nodos ou neurônios), maciçamente paralelos e com capacidade de generalização, ou seja, podem produzir resultados satisfatórios para dados não fornecidos no momento do aprendizado da rede, Jr e Montgomery (2007).

O neurônio matemático, assim como o biológico, recebe um ou mais sinais de entrada e devolve um único sinal de saída. Esse sinal de saída pode se tornar o sinal de entrada para um ou mais neurônios da próxima camada. Como os sinais chegam simultaneamente aos neurônios, a rede é classificada como sistemas de

processamento paralelo. A intensidade de cada ligação é representada por uma grandeza chamada peso sináptico.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Séries Temporais

2.1.1 Definição e Características

Ao conjunto de observações aleatórias $X_1, X_2, X_3, \dots, X_t$, observadas em instantes discretos de tempo igualmente espaçados 1, 2, 3, ..., t , dá-se o nome de série temporal, Cordeiro (2002).

As principais características de uma série temporal são:

- Existe uma dependência entre os dados vizinhos.
- Interesse na análise e modelagem da dependência entre os dados.
- A ordem dos dados é fundamental. Uma característica muito importante deste tipo de dados é que as observações vizinhas são dependentes e o interesse é analisar e modelar esta dependência.
- Diferentemente de outros procedimentos estatísticos, nas séries temporais faz-se necessário o uso de procedimentos específicos, devido à dependência entre os dados.

Segundo Ehlers (2009) as séries temporais podem apresentar as seguintes características:

Sazonalidade - quando uma série temporal apresenta um comportamento que tende a se repetir a cada n intervalos de tempo, chama-se a essa característica de sazonalidade. Por exemplo, espera-se um aumento acentuado nas vendas, no comércio em geral, no mês de dezembro e isso tende a se repetir ao longo dos anos. Se esse aumento for aproximadamente constante independente do nível global da série, tem-se uma sazonalidade aditiva, caso esse aumento dependa do nível global da série, temos uma sazonalidade multiplicativa.

Tendência - globalmente, uma série pode exibir tendência de crescimento (ou decréscimo) com vários possíveis padrões.

- Crescimento linear - quando essa tendência é praticamente fixa, não dependendo do valor global da série. Por exemplo, a cada ano o aumento esperado nas vendas de certo brinquedo é de um milhão de reais.
- Crescimento exponencial - se essa tendência depender do valor global da série. Por exemplo, a cada ano as vendas de certo brinquedo aumentam de um fator 1,3.
- Crescimento amortecido - quando a tendência depender da tendência anterior. Por exemplo, as vendas de certo brinquedo tem um aumento esperado de 70% sobre o ano anterior. Se o aumento esperado for de um milhão de reais no primeiro ano, no segundo ano será de 700 mil reais, no terceiro ano serão de 490 mil reais e assim por diante.

2.1.2 Previsão de séries temporais

Dada uma série temporal $Z_T = \{X_1, X_2, X_3, \dots, X_t\}$ podemos fazer sua previsão k instantes à frente Z_{T+k} . O número de instantes à frente para o qual é feita a previsão é chamado de horizonte de previsão, Barros (2007).

Um método usado para se fazer uma previsão pode ser definido como sendo um conjunto de procedimentos utilizados na previsão de uma série temporal. Analisaremos adiante três métodos que podem ser usados na previsão de uma série temporal, porém vale salientar que um método não é necessariamente melhor que outro, e sim que um método será mais adequado que outro dependendo das características da série.

Método ingênuo - neste método a previsão de uma observação consiste na observação anterior, ou seja, $Z_{T+1} = Z_T$, onde Z_{T+1} representa o valor previsto e Z_T representa o valor anterior.

Um exemplo clássico é a previsão do preço de uma ação - geralmente a melhor previsão para o preço de amanhã é o preço de hoje, Barros (2007).

Média móvel - para prever um determinado valor usando esse método devemos calcular a média aritmética dos n valores anteriores. Matematicamente podemos representar da seguinte forma:

$$Z_{T+1} = \frac{Z_T + Z_{T-1} + Z_{T-2} + Z_{T-3} + \dots + Z_{T-(n-1)}}{n} \quad [\text{Eq.01}]$$

Na Eq.01, n representa a janela de previsão a ser usada. Quanto maior o valor de n mais suave tende a ser a previsão, por outro lado quanto menor a janela de previsão mais esse valor tende a oscilar, Barros (2007).

Esse método recebe o nome de média móvel, pois assim que a próxima observação se torna disponível a média das observações é recalculada desprezando o valor mais antigo, Mueller (1996). Uma característica importante nesse método é que todas as observações possuem os mesmos pesos, já que é extraída uma média aritmética das observações.

Amortização exponencial - esse método é muito parecido com o da média móvel, porém nesse método cada observação terá um peso. Esse peso vai diminuindo para os valores mais antigos, ou seja, os valores mais recentes influenciam mais no resultado da previsão. Podemos representar esse modelo matematicamente da seguinte forma:

$$Z_{T+1} = \alpha Z_T + \alpha(1 - \alpha)Z_{T-1} + \alpha(1 - \alpha)^2 Z_{T-2} + \dots + \alpha(1 - \alpha)^{T-1} Z_{T-(n-1)} \quad [\text{Eq.02}]$$

Onde Z_{T+1} representa o valor previsto um instante à frente, Z_T é a observação anterior, α é o coeficiente de amortização e n representa os n termos da série temporal.

2.2 Redes Neurais Artificiais

Sistemas baseados em RNAs são compostos por unidades de processamento (nodos ou neurônios), maciçamente paralelos e com capacidade de generalização, ou seja, podem produzir resultados satisfatórios para dados não fornecidos no momento do aprendizado da rede, Jr e Montgomery (2007).

Na figura 1 temos o modelo do neurônio matemático. Esse neurônio, assim como o biológico, recebe um ou mais sinais de entrada e devolve um único sinal de saída. Esse sinal de saída pode se tornar o sinal de entrada para um ou mais neurônios da próxima camada. Como os sinais chegam simultaneamente aos neurônios, a rede é classificada como sistemas de processamento paralelo. A intensidade de cada ligação é representada por uma grandeza chamada peso sináptico.

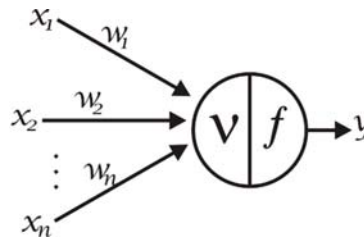


Figura 1 – Neurônio matemático

Quando as entradas, x_n , são informadas ao neurônio, são então multiplicadas pelos seus respectivos pesos sinápticos, w_n . O neurônio, então, efetua o somatório desses produtos.

$$v = \sum_{i=1}^n w_i x_i \quad [\text{Eq.03}]$$

A equação Eq.03 é denominada função de combinação ou função de ativação. O resultado deste somatório é, então, apresentado para a função f , função de transferência (ver Eq.04 e Eq.05), que entre suas funções pode ser citada sua capacidade de evitar o acúmulo progressivo dos valores de saída ao longo das camadas da rede, já que tais funções possuem valores máximos e mínimos em intervalos bem determinados, Jr e Montgomery (2007).

$$f = \frac{1}{1 + e^{-v}} \quad [\text{Eq.04}]$$

$$f = e^{-v^2} \quad [\text{Eq.05}]$$

A equação Eq.04 é chamada Função Sigmóide e a equação Eq.05 é a Função Gaussiana. São algumas das funções de transferência usadas.

Uma RNA pode possuir uma ou várias camadas. Dependendo da conexão entre camadas podem-se ter várias estruturas diferentes. A forma como estas conexões estão dispostas está intimamente relacionada com o algoritmo de aprendizagem usado para o treinamento da rede. Uma rede feedforward é aquela em que a informação segue apenas um sentido, ou seja, a informação vai passando de uma camada para outra até chegar à saída. Já uma rede com recorrência, a informação que entra em um neurônio pode vir de neurônios da camada seguinte, ou até mesmo do próprio neurônio.

2.3 Reservoir Computing

As redes neurais recorrentes (RNR), ver figura 2, representam uma grande e variada classe de modelos computacionais baseados no funcionamento da rede neural biológica. Echo State Network (ESN) e Liquid State Machine (LSM) introduziram um novo paradigma de redes neurais recorrentes, conhecido como Reservoir Computing (RC), que oferece uma metodologia intuitiva que reduz o esforço para treiná-la, Lukosevicius e Jaeger (2009). A diferença entre as ESNs e as LSMs é que a segunda teve uma inspiração baseada na realidade biológica enquanto a primeira foi desenvolvida para resolver problemas matemáticos.

O fato de essas redes possuírem recorrência causa um grande impacto em seu funcionamento, tais como:

- Desenvolvimento de uma ativação dinâmica temporal autossustentada em seus caminhos interconectados, mesmo na ausência de sinais de entrada, Lukosevicius e Jaeger (2009).
- Quando sinais de entrada são dirigidos à rede, ela preserva no seu estado interno uma transformação não linear do histórico das entradas, ou seja, há uma memória dinâmica, e então a rede estará apta a processar informação de contexto temporal, Lukosevicius e Jaeger (2009).

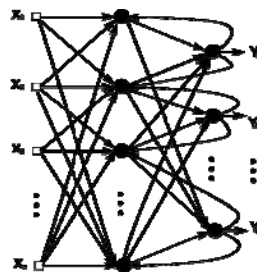


Figura 2 – Rede neural recorrente

Neste trabalho usamos as ESN na criação do sistema.

3 DESCRIÇÃO DA PROPOSTA

O objetivo desse trabalho é a apresentação dos resultados da criação de um sistema inteligente híbrido para previsão de séries temporais, através de um estudo de caso da previsão em curto prazo da velocidade média horária dos ventos para as cidades de Belo Jardim, Triunfo e São João do Cariri no estado de Pernambuco. O sistema inteligente híbrido é composto por Rede Neural Artificial e Algoritmo Genético (AG). Este trabalho apresenta também os resultados obtidos com a utilização dos métodos estatísticos de amortização exponencial e média móvel, para previsão da mesma série temporal (ver tabela 1).

Tabela 1 – Bases de dados

Cidade	Período de Observação	Número de Registros
Belo Jardim (BJD)	01/07/2004 a 31/12/2005	13.176
Triunfo	01/07/2004 a 31/12/2006	21.936
São João do Cariri	01/01/2006 a 31/12/2007	17.520

4 BASE DE DADOS, METODOLOGIA E ESTUDO DE CASO

4.1 Base de dados

Os dados utilizados no nosso trabalho são dados disponibilizados pelo projeto SONDA (Sistema de Organização Nacional de Dados Ambientais). O projeto SONDA é um projeto do Instituto Nacional de Pesquisas Espaciais (INPE) para implementação de uma infraestrutura física e de recursos humanos destinados ao levantamento e melhoramento do banco de dados dos recursos de energia solar e eólica no Brasil.

Escolhemos as séries de velocidade dos ventos de Belo Jardim (BJD), Triunfo (TRI) e São João do Cariri (SCR), que são cidades do estado de Pernambuco e estão respectivamente a 718m, 718m e 1123m de altitude. Os dados utilizados nos nossos experimentos correspondem às velocidades médias horárias exibidos na Tabela 1, vista no tópico anterior.

4.2 Metodologia

Para resolver um problema de aprendizado de pesos de uma rede neural usando AGs, é necessário transformar a estrutura do problema em alguma forma compreensível para o AG, criando uma codificação adequada. Tem-se, pois, que desenvolver uma estrutura para o AG tal que se possa restringir em blocos a influência da alteração de sinapses, de modo que os operadores de crossover e de mutação possam operar e gerar resultados consistentes.

Em geral, o treinamento das redes neurais artificiais é realizado através de repetidas tentativas com diferentes topologias de rede, até serem obtidos resultados satisfatórios para o problema. Além de consumir tempo e exigir experiência no treinamento de redes neurais artificiais, este processo pode obter redes com conexões e unidades de processamento desnecessárias. Sem considerar o fato de que quanto maior a topologia da rede, mais complexo é o ajuste do valor destas conexões. Desta forma, a otimização de topologia e pesos de redes neurais artificiais através de algoritmos evolucionários é uma abordagem interessante para a geração de redes eficientes com topologias pequenas.

Na metodologia proposta neste trabalho, cada solução é codificada por um vetor que é formado na seguinte ordem por:

1. Número de nodos no reservóir;
2. Se há ou não conexão entre a entrada e a saída da rede;
3. Se há ou não conexão entre o bias e a saída;
4. Se há ou não conexão recorrente entre a saída com ela própria;
5. Se há ou não conexão entre o bias e o reservóir;
6. Se há ou não conexão entre saída e o reservóir;

7. Função do nodo (0-tangente hiperbólica ou 1-sigmoidal);
8. Tipo do treinamento (0-Pseudo inversa ou 1-Ridge-regress);
9. Leak rate, que varia entre [0.1;1]
10. Valor do parâmetro de regularização, que varia entre $[10^{-8}; 10^{-1}]$;
11. É um vetor que tem os valores dos pesos entre as ligações entre reservóir com ela mesma, bias com reservóir e saída com o reservóir. O peso dos valores varia entre [-1; 1].

A população (solução) inicial é gerada aleatoriamente com 120 indivíduos. A nova solução é gerada a partir de uma simulação da reprodução sexuada através dos operadores genéticos dos pais que foram escolhidos através do método de seleção chamado de Método da Amostragem Estocástica Uniforme de acordo com sua avaliação. A avaliação é dada pela Eq. 06, sendo tão melhor quanto menor for o erro. Este erro é obtido comparando-se a saída obtida com a saída desejada através do MSE (Erro Médio Quadrático – Eq. 07).

$$fitness = MSE_{treinamento} + |MSE_{treinamento} - MSE_{validação}| \quad [Eq.06]$$

$$MSE_{\%} = 100 \times \frac{L_{max} - L_{min}}{N \cdot P} \sum_{p=1}^P \sum_{i=1}^N (L_{pi} - T_{pi})^2 \quad [Eq.07]$$

onde L_{max} e L_{min} são os valores máximo e mínimo das velocidades dos ventos, respectivamente; N é o número de unidades de saída da RNR. P é o número total de padrões na base de dados; L_{pi} e T_{pi} são a saída atual e desejada do i -ésimo neurônio da camada de saída, respectivamente.

O elitismo foi usado com o valor de 2. O que significa que 2 indivíduos dos 120 da população com as melhores avaliações serão escolhidos para sobreviver para a próxima geração. Com o elitismo pode haver perda de diversidade na população, muito embora se garanta que as possíveis melhores características das soluções não sejam perdidas.

Para comparar com a previsão feita usando-se rede neural, foi utilizado dois métodos estatísticos, média móvel e amortização exponencial.

4.3 Estudo de Caso

A metodologia descrita acima foi aplicada à busca da melhor ESN para previsão da velocidade média horária dos ventos de Belo Jardim (BJD). A figura 3a apresenta a topologia de uma ESN na forma de matriz, onde um quadrado em branco representa a existência de conexão e um quadrado preto significa que não há conexão. A figura 3b apresenta os parâmetros globais da melhor ESN: número de nodos no reservóir, função de ativação dos neurônios, função de treinamento da camada de readout, leak rate e parâmetro de regularização. Após 11 gerações, a configuração da melhor ESN escolhida pelo método proposto apresentou a topologia e os parâmetros mostrados na Figura 3. A ESN possui todas as possíveis conexões, ou seja, conexões entre bias (B) e reservóir (R), entrada (E) e reservóir, entrada e saída (S), conexão recorrente no reservóir, conexão entre saída e reservóir e conexão recorrente na camada de saída.



(a)

Nodos no Reservóir	52
Função de ativação	Tangente Hiperbólica
Função de Treinamento	Ridge Regress
Leak Rate	0.4
Parâmetro de regularização	0.0001

(b)

Figura 3 – Configuração da topologia da melhor ESN após 11 gerações. Quadrado branco significa que há conexão entre as partes correspondentes e quadrado preto que não existe.

A figura 4 apresenta a evolução do melhor fitness ao longo das 11 gerações. Percebe-se que após a geração 4 o algoritmo genético se estabiliza perto do melhor fitness encontrado 0.0053995. O fitness é baseado no MSE do conjunto de treinamento e do conjunto de validação. Então, quanto menor o fitness, menor o erro desta rede para previsão da série em questão. Ver Eq. 06 e Eq. 07.

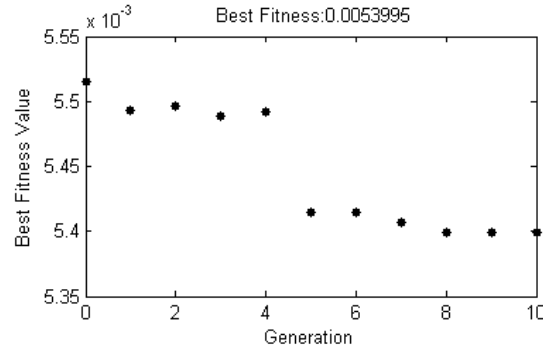


Figura 4 – Evolução do melhor fitness ao longo das 11 gerações.

Na figura 5 temos a previsão através dos métodos estatísticos média móvel e amortização exponencial. Percebe-se que os resultados da previsão dos dois métodos são muito similares para essa série.

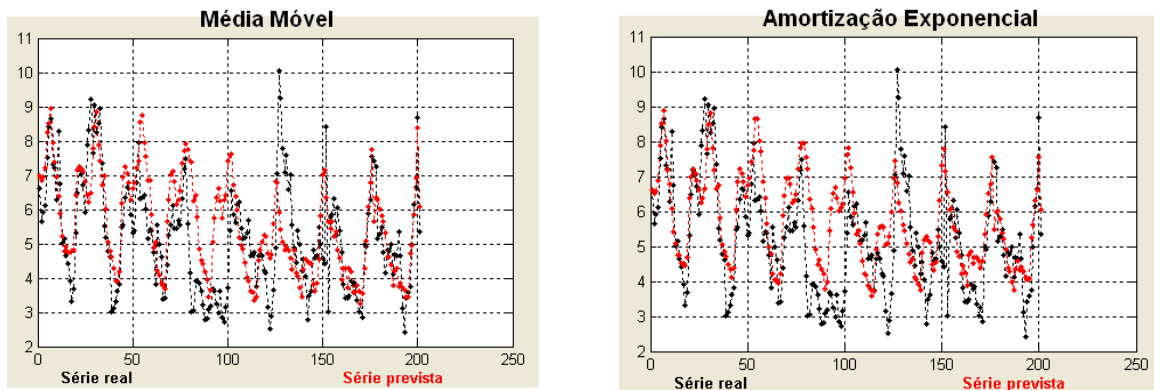


Figura 5 – Previsão com Média Móvel e Amortização Exponencial

5 DISCUSSÃO E CONSIDERAÇÕES FINAIS

Na tabela 2 temos o MSE, MAPE e MAE para previsão realizada pelo sistema inteligente híbrido (RNA e AG), para média móvel e para amortização exponencial. Quando comparamos os resultados sistema inteligente híbrido com os dois métodos estatísticos fica claro que, para a série temporal usada nesse trabalho, os melhores resultados foram os obtidos pelo sistema inteligente híbrido, tanto no treinamento como no teste. Se compararmos os resultados apenas entre os métodos estatísticos, perceber-se-ia que não houve grande diferença entre as previsões realizadas pelo método de média móvel e o método de amortização exponencial. Porém, vale salientar que a previsão de uma série temporal pode ser muito diferente, no que diz respeito à sua precisão, quando comparada a outra série para o mesmo método de previsão. Em outras palavras, um método pode ser muito preciso em sua previsão para uma série e não ser tão preciso para a previsão de outra.

Os métodos estatísticos de média móvel e de amortização exponencial são muito mais precisos para séries temporais com curva de crescimento ou decrescimento sem grandes oscilações, ou seja, quando os valores

oscilam muito fortemente o erro na previsão acaba sendo maior. Já a rede neural artificial consegue se adaptar melhor a essas variações.

Tabela 2 – Comparação entre os resultados do Sistema Inteligente Híbrido com os Métodos Estatísticos para a cidade de Belo Jardim

	Sistema Inteligente Híbrido			Média Móvel			Amortização Exponencial		
	MSE	MAPE (%)	MAE (m/s)	MSE	MAPE (%)	MAE (m/s)	MSE	MAPE (%)	MAE (m/s)
Treinamento	0,5001	17,49	0,6946	2,1198	29,95	1,1557	1,9793	27,68	1,1004
Teste	0,394	12,1	0,6221	2,1318	29,01	1,1497	1,8643	27,56	1,073

Tabela 3 – Comparação entre os resultados do Sistema Inteligente Híbrido com os Métodos Estatísticos para a cidade de Triunfo

	Sistema Inteligente Híbrido			Média Móvel			Amortização Exponencial		
	MSE	MAPE (%)	MAE (m/s)	MSE	MAPE (%)	MAE (m/s)	MSE	MAPE (%)	MAE (m/s)
Treinamento	0,08769	8,435	0,835	6,0847	24,6672	1,88395	5,49351	18,1518	1,79306
Teste	0,09945	9,88	0,8706	5,4446	23,0163	1,81731	5,049	21,9945	1,73792

Tabela 4 – Comparação entre os resultados do Sistema Inteligente Híbrido com os Métodos Estatísticos para a cidade de São João do Cariri

	Sistema Inteligente Híbrido			Média Móvel			Amortização Exponencial		
	MSE	MAPE (%)	MAE (m/s)	MSE	MAPE (%)	MAE (m/s)	MSE	MAPE (%)	MAE (m/s)
Treinamento	0,3766	17,39	0,674	2,08095	28,5424	0,285424	1,87618	27,3392	1,05958
Teste	0,3265	13,36	0,6417	2,53976	36,5232	1,23298	2,36957	34,7901	1,18024

6 AGRADECIMENTOS

Agradeço ao CNPq por financiar esse trabalho.

7 REFERÊNCIAS BIBLIOGRÁFICAS

BARROS, M. **Processos Estocásticos**. Publit, 2007.

BRAGA, A. D.; CARVALHO, A. P.; LUDERMIR, T. B. **Redes Neurais Artificiais - Teoria e Aplicações**. Rio de Janeiro: LTC Editora, 2007.

CORDEIRO, D. M. **Séries Temporais - Análise Quantitativa - Teoria e Aplicações**. Recife: EDUPE, 2002.

EHLERS, R. S. **Análise de Séries Temporais**. Disponível em <<http://www.icmc.usp.br/~ehlers/notas/stemp.pdf>>, acesso em 01 de Julho de 2009.

JR, O. L.; MONTGOMERY, E. **Redes Neurais - Fundamentos e Aplicações com Programas e C**. Rio de Janeiro: Ciência Moderna Ltda, 2007.

LUKOSEVICIUS, M.; JAEGER, H. **Reservoir computing approaches to recurrent neural network training**, Computer Science Review 3, pp.127-149, 2009.

MUELLER, A. **Uma aplicação de Redes Neurais Artificiais na Previsão do Mercado**. Dissertação de mestrado, Universidade Federal de Santa Catarina, disponível em <<http://www.eps.ufsc.br/disserta96/mueller/index/index.htm>>, Julho de 1996.