

Composição de Redes Neurais Artificiais com Algoritmo Genético Modificado para a Previsão de Séries Temporais

Paulo S. G. de Mattos Neto, Gustavo G. Petry, João P. de M. Atide e Tiago A. E. Ferreria

Resumo—Neste trabalho é apresentada uma abordagem híbrida com a composição de um algoritmo genético modificado e redes neurais artificiais do tipo MLP com três formas distintas de modelagens. A metodologia apresentada exhibe uma forma relativamente simples de se montar experimentos baseados em redes neurais quando comparada à forma convencional de escolha e ajustes dos parâmetros de uma rede. Em especial, a metodologia desenvolvida é aplicada ao problema de previsão de séries temporais, apresentando resultados experimentais com o uso de um conjunto de quatro séries relevantes do mundo real.

Index Terms—Algoritmos Genéticos, Redes Neurais, Séries Temporais.

I. INTRODUÇÃO

O ramo da ciência que naturalmente aborda o problema de previsão de séries temporais é a estatística. Na prática, são encontradas várias referências aos modelos ARIMA [1] para a previsão de séries temporais. Contudo tais modelos são lineares, o que pode vir a ser um limitador.

Grandes esforços têm sido feitos para o desenvolvimento de modelos estatísticos não-lineares, entretanto um fator complicante para tais modelos vem da alta complexidade matemática. Por esta razão, um novo e promissor conceito de não linearidade vem sendo muito explorado nos últimos anos, as Redes Neurais Artificiais (RNA) [2]. Propostas baseadas em RNA's vêm sendo utilizadas como uma forma de modelagem não-linear de séries temporais [3], contudo, as RNA's requerem ajustes em um vasto conjunto de parâmetros. Em teoria, o estabelecimento de todos estes parâmetros poderia necessitar a investigação de todo o espaço de estados, e desde que este fato é impraticável, pode-se utilizar algum tipo de algoritmo de busca, como por exemplo os Algoritmos Genéticos (AG) [4], para encontrar uma solução sub-ótima a qual tenha um desempenho satisfatório a um custo computacional aceitável.

Neste trabalho é exposto um sistema híbrido inteligente com a utilização de um AG modificado e RNA's como uma abordagem ao problema de previsão de séries temporais, que efetivamente busca: 1. os específicos retardos temporais relevantes (*lags*) da série temporal dentro de um espaço de busca pré-determinado pelo usuário, 2. a melhor arquitetura da RNA através de um processo de poda [5] e 3. o valor dos pesos das conexões da RNA. Um conjunto de sete medidas de desempenho são utilizadas para a análise dos experimentos realizados com um grupo de

quatro séries relevantes do mundo real.

II. PROBLEMA DE PREVISÃO

Uma série temporal é um conjunto de pontos, geralmente equidistantes no tempo, tal como, $Z_t = \{z_t \in \mathbb{R} \mid t = 1, 2, 3 \dots N\}$, onde t é um índice temporal e N é o número de observações. O objetivo de se aplicar técnicas de previsão para uma dada série temporal Z_t é identificar padrões presentes no conjunto de pontos da série, criando um modelo capaz de construir os próximos padrões temporais.

Os relacionamentos entre os dados históricos da série temporal constituem um espaço de fase d -dimensional. Desta maneira, um espaço de fase d -dimensional pode ser construído de tal forma que contenha toda a informação referente a uma série temporal de interesse. F. Takens [6] provou que se d é suficientemente grande, tal espaço construído é homeomórfico ao espaço de fase que originalmente gerou a série. O teorema de Takens é então a justificativa teórica da possibilidade de reconstrução do espaço de fase utilizando os retardos temporais, e se tal espaço é construído corretamente, o teorema de Takens também garante que a dinâmica do espaço reconstruído é idêntica a dinâmica do espaço real.

O grande problema na reconstrução do espaço de fase é naturalmente a escolha da dimensionalidade d , ou mais precisamente, a escolha dos retardos temporais relevantes necessários para caracterização da dinâmica do sistema. Desta forma, a metodologia exposta é baseada em um processo de busca pelos retardos temporais relevantes para a correta caracterização do processo que governa a série temporal em estudo.

III. MÉTODO PROPOSTO

A possibilidade de melhora no desempenho de predição das RNA's pode ser alcançada com o correto ajuste dos seus parâmetros. O grande problema em questão é como determinar o valor destes parâmetros, sendo estes dependentes do problema e formando um vasto universo de possibilidades. Uma possível solução é a aplicação de alguma técnica de busca que seja capaz de realizar a escolha destes parâmetros da melhor maneira possível.

A idéia aplicada neste trabalho é o desenvolvimento de um sistema híbrido inteligente com a combinação de um AG modificado e RNA's. Cada indivíduo da população do AG representa uma RNA do tipo MLP de três camadas onde o cromossomo é montado com todos os parâmetros a serem ajustados: pesos e conexões. A quantidade máxima de entradas (*lags*) e de neurônios in-

Departamento de Estatística e Informática - UNICAP, Rua do Príncipe, 526, Boa Vista, 50500-900, Recife, PE, Brasil. E-mail: {paulosalgado10, gugapetry, jpataide}@gmail.com, taef@dei.unicap.br

termediários são constantes definidas pelo usuário, já a camada de saída será constituída apenas por um neurônio (previsão de um passo). A Figura 1 mostra um esquema genérico para uma RNA utilizada. Em cada conexão existe uma chave liga/desliga, representada por um pequeno quadrado. Estas chaves podem ativar ou desativar as respectivas conexões — procedimento de poda [5]. Este processo de poda dá a capacidade efetiva ao sistema de realizar uma busca pela quantidade mínima necessária de retardos temporais relevantes, pela quantidade mínima de neurônios da camada escondida, e assim, pela rede mais compacta possível.

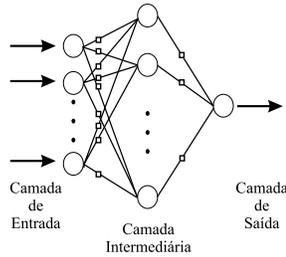


Figura 1. Esquema genérico da rede neural artificial codificada como indivíduo da população do algoritmo genético.

Os indivíduos da população são avaliados a partir da função de *fitness*,

$$Fitness = \frac{1}{1 + MSE} \quad (1)$$

onde *MSE* é o erro médio quadrático de previsão da RNA.

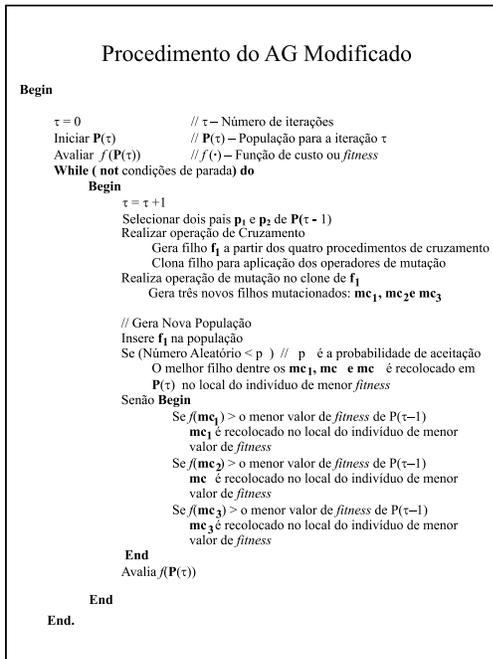


Figura 2. Procedimento do Algoritmo Genético modificado utilizado nos experimentos.

Desta forma, o AG modificado, como descrito na Figura 2, utiliza operadores de cruzamento e mutação modificados

com o intuito de acelerar a convergência para uma possível solução (a melhor RNA possível), como mostrado no trabalho de Leung et al [4]. São aplicados quatro diferentes operadores de cruzamento, gerando quatro novos cromossomos: C_1 , C_2 , C_3 e C_4 , definidos por,

$$C_1 = [C_1^1 C_2^1 \dots C_{nVar}^1] = \frac{P_1 + P_2}{2} \quad (2)$$

$$C_2 = [C_1^2 C_2^2 \dots C_{nVar}^2] = P_{max}(1 - w) + max(P_1, P_2) \quad (3)$$

$$C_3 = [C_1^3 C_2^3 \dots C_{nVar}^3] = P_{min}(1 - w) + min(P_1, P_2) \quad (4)$$

$$C_4 = [C_1^4 C_2^4 \dots C_{nVar}^4] = \frac{(P_{max} + P_{min})(1 - w) + (P_1 + P_2)w}{2} \quad (5)$$

onde P_1 e P_2 são os cromossomos pais escolhidos com uma chance proporcional aos seus valores de *fitness*, $P_{max} = [P_{max}^1 P_{max}^2 \dots P_{max}^{nVar}]$ e $P_{min} = [P_{min}^1 P_{min}^2 \dots P_{min}^{nVar}]$ são os valores máximos e mínimos dos genes. O peso $w \in [0, 1]$ é definido pelo usuário, onde quanto maior seu valor maior será a contribuição direta dos pais. E por fim, as funções $max(P_1, P_2)$ e $min(P_1, P_2)$ irão escolher o genes de maior e menor valor respectivamente. As operações de cruzamento dadas pela Equação (2) e pela Equação (5) tendem a varrer a região central do domínio de busca (se $w \approx 1$ na Equação (5)), e as Equações (3) e (4) tendem a mover a busca para os limites máximo e mínimo do domínio respectivamente.

Dentre os quatro cromossomos gerados, aquele que detiver o maior valor de *fitness* será eleito como o filho gerado pelo processo de cruzamento.

Depois da seleção do filho, este é clonado. Seu clone irá sofrer um processo de mutação, e três novos filhos serão criados. Sendo o clone do filho eleito no processo de cruzamento dado por C_g ($g = 1, 2, 3, 4$), gera-se as mutações da forma:

$$MC_j = [C_1^g C_2^g \dots C_{nVar}^g] + [\gamma_1 \Delta MC_1 \gamma_2 \Delta MC_2 \dots \gamma_{nVar} \Delta MC_{nVar}] \quad (6)$$

onde γ_i ($i = 1, 2, \dots, nVar$) assume o valor 0 ou o valor 1, e ΔMC_i ($i = 1, 2, \dots, nVar$) são números aleatórios gerados de tal forma a satisfazerem a condição: $P_{min}^i \leq C_i^g + \Delta MC_i \leq P_{max}^i$ e $nVar$ é a quantidade de genes por cromossomo.

A primeira mutação ($j = 1 \Rightarrow MC_1$) é realizada com apenas um único γ_i igual a 1 (escolhido aleatoriamente), e os demais iguais a zero. A segunda mutação ($j = 2 \Rightarrow MC_2$) decorre com uma quantidade aleatória de γ_i 's (escolhidos ao acaso) iguais a 1 e os demais iguais a zero. E, por fim, a terceira mutação ($j = 3 \Rightarrow MC_3$) ocorre com todos os γ_i 's iguais a 1.

Cada um dos filhos mutacionados serão aceitos se: 1. o seu *fitness* for superior ao *fitness* do pior indivíduo da população; ou 2. por sorteio, segundo uma probabilidade de aceitação de mutação (P_{Mut}).

A. Modelagem dos Indivíduos

Como já explicitado, cada indivíduo da população do AG é uma codificação para uma RNA do tipo MLP com três camadas. Estes indivíduos são representados por cromossomos que contem os seguintes genes(parâmetros):

- S_{ij}^1 : chaves das conexões da camada de entrada para a camada intermediária;
- S_{jk}^2 : chaves das conexões da camada intermediária para a camada de saída;
- S_j^1 : chaves das conexões dos bias para os neurônios da camada intermediária;
- S_k^2 : chave da conexão do bias para o neurônio da camada de saída;
- W_{ij} : pesos das conexões da camada de entrada para a camada intermediária;
- W_{jk} : pesos das conexões da camada intermediária para a camada de saída;
- b_j^1 : bias dos neurônios intermediários;
- b_k^2 : bias do neurônio de saída

podendo todos estes parâmetros assumirem valores reais.

Desta forma, são propostas três formas distintas de modelagem para as RNAs.

A.1 Modelo REDE 1

A primeira configuração para a modelagem das RNAs utiliza neurônios na camada intermediária com função de ativação sigmóide logística e os neurônios da camada de saída com função de ativação linear e com os valores de bias passando por uma função sigmóide logística. A resposta desta rede, para um neurônio de saída k é dada pela Equação 7,

$$y_k(t) = \sum_{j=1}^{n_h} \delta(S_{jk}^2) W_{jk} \text{Sig} \left[\sum_{i=1}^{n_{in}} (\delta(S_{ij}^1) W_{ij} Z_i(t) - \delta(S_j^1) b_j^1) \right] - \delta(S_k^2) \text{Sig}(b_k^2) \quad (7)$$

onde Sig é a função sigmóide logística e $\delta(x)$ é uma função degrau, se $x < 0 \Rightarrow \delta(x) = 0$ e se $x \geq 0 \Rightarrow \delta(x) = 1$. $Z_i(t)$ ($i = 1, 2, \dots, n_{in}$) são os valores de entrada da rede, n_{in} denota o número de entradas e n_h é o número de neurônios escondidos. Para o caso particular de previsão de um passo, tem-se apenas um único neurônio de saída, ou seja, $k = 1$.

A.2 Modelo REDE 2

A segunda configuração foi considerada como uma rede com a camada intermediária utilizando a sigmóide logística como função de ativação, e a camada de saída a função linear sem a passagem dos valores dos seus bias por uma função sigmóide logística, resultando a saída da rede como explicitado pela Equação 8,

$$y_k(t) = \sum_{j=1}^{n_h} \delta(S_{jk}^2) W_{jk} \text{Sig} \left[\sum_{i=1}^{n_{in}} (\delta(S_{ij}^1) W_{ij} Z_i(t) - \delta(S_j^1) b_j^1) \right] - \delta(S_k^2) (b_k^2) \quad (8)$$

A.3 Modelo REDE 3

A terceira opção considerada para a modelagem dos indivíduos foi uma rede com todos os neurônios com funções

sigmóide logísticas como funções de ativação, resultando a saída da rede dada pela Equação 9,

$$y_k(t) = \text{Sig} \left\{ \sum_{j=1}^{n_h} \delta(S_{jk}^2) W_{jk} \text{Sig} \left[\sum_{i=1}^{n_{in}} (\delta(S_{ij}^1) W_{ij} Z_i(t) - \delta(S_j^1) b_j^1) \right] - \delta(S_k^2) (b_k^2) \right\} \quad (9)$$

B. Medidas de Desempenho

Na previsão de séries temporais, a medida de desempenho mais intuitiva é o próprio erro de previsão: o valor atual da série (*target*) menos a resposta do modelo (*output*),

$$e_t = (\text{target} - \text{output}) \quad (10)$$

A forma mais popular da utilização de e_t é o Erro Médio Quadrático, definido por,

$$MSE = \frac{\sum (e_t)^2}{N} \quad (11)$$

onde N é a quantidade de pontos previstos da série.

Embora o erro MSE possa ser utilizado como parâmetro da função de *fitness*, este não exhibe, de forma clara, o comportamento do modelo de previsão, gerando uma polêmica quanto ao seu uso [7]. Desta forma, serão utilizadas outras medidas de desempenho como suporte para a análise dos dados, gerando uma base mais consolidada de resultados.

Dentre as medidas adotadas, tem-se Erro Médio Percentual ($MAPE$),

$$MAPE = \frac{1}{N} \sum \left| \frac{e_t}{Z_t} \right| \quad (12)$$

onde N é a quantidade de pontos previstos, e_t é dado pela Equação 10 e Z_t é o ponto da série no instante t .

Outra medida de desempenho adotada é a estatística U de Theil (ou $NMSE$), que se baseia no MSE do preditor, normalizado pelo erro de previsão supondo-se um modelo do tipo *Random Walk*. Um modelo *Random Walk* simplesmente assume que o melhor valor para o tempo $t + 1$ é o valor obtido no tempo t , mais um termo de ruído. A estatística U de Theil é definida por,

$$\text{Theil} = \frac{\sum (\text{target}_t - \text{output}_t)^2}{\sum (\text{output}_t - \text{output}_{t-1})^2} \quad (13)$$

onde se $\text{Theil} = 1$ o modelo tem um desempenho igual a um modelo do tipo *Random Walk*, se $\text{Theil} > 1$ o desempenho é inferior a um modelo do tipo *Random Walk* e se $\text{Theil} < 1$ o desempenho é superior a um modelo do tipo *Random Walk*.

Outra medida de desempenho vem da precisão de se prever as alterações da direção, ou seja, se o valor futuro irá subir ou cair em relação ao valor atual. A medida $POCID$ mede o percentual de acerto quanto a tendência da série, definida por,

$$POCID = 100 \frac{\sum_{\mu=1}^N D_{\mu}}{N} \quad (14)$$

Onde,

$$D_\mu = \begin{cases} 1, & \text{se } (Target_\mu - Target_{\mu-1})(Output_\mu - Output_{\mu-1}) > 0 \\ 0, & \text{caso contrário} \end{cases} \quad (15)$$

Uma medida bastante interessante é a ARV, dada pela Equação 16,

$$ARV = \frac{\sum_{i=1}^N (output_i - target_i)^2}{\sum_{i=1}^N (output_i - \overline{target})^2} \quad (16)$$

onde \overline{target} é a média da série. Se $ARV = 1$, o modelo de previsão tem um desempenho idêntico a se prever simplesmente a média da série, se $ARV > 1$ o modelo de previsão é inferior a se prever a média da série. Desta forma, para o modelo de previsão ser eficiente, quando comparado a previsão da média da série, deve-se ter $ARV < 1$.

Todas estas medidas apresentadas até então não levam em consideração a quantidade de parâmetros de ajuste que o modelo contém. Quanto maior a quantidade de parâmetros livres maior a possibilidade de superestimação (ou *overfitting*). Os critérios de informação *AIC* (*Akaike Information Criterium*) e *BIC* (*Baysean Information Criterium*) levam em consideração a quantidade de parâmetros livres do sistema, penalizando o modelo proporcionalmente à sua quantidade de parâmetros livres. Estes critérios podem ser definidos de forma aproximada como [8],

$$AIC = N \ln(MSE) + 2p \quad (17)$$

$$BIC = N \ln(MSE) + p + p \ln(N) \quad (18)$$

onde N é a quantidade de pontos previstos e p denota o número de parâmetros livres do modelo.

IV. RESULTADOS EXPERIMENTAIS

Antes da execução dos experimentos foi realizado um processo de normalização dos dados de tal forma que todos os pontos das séries temporais ficassem contidos no intervalo $[0,1]$. Depois da normalização, as séries foram divididas em três conjuntos: conjunto de treinamento (50% dos dados); conjunto de validação (25% dos dados); e conjunto de teste (25% dos dados).

Em todos os experimentos foram utilizados sempre a mesma arquitetura inicial para as RNAs, sendo uma rede 10-10-1 totalmente conectada, ou seja, 10 entradas, 10 neurônios na camada intermediária e 1 neurônio na camada de saída com dez inicializações distintas do sistema para cada modelo de RNA. Também foram utilizados os mesmo valores para os parâmetros do algoritmo para todas as séries: Probabilidade de Aceite de Mutação (P_{Mut}) = 0.1; Peso usado nos operados de cruzamento (w) = 0.9; Quantidade de Iterações = 10000.

O sistema híbrido implementou a validação cruzada e utilizou como critério de parada 1. a quantidade máxima de iterações, 2. taxa de crescimento do *fitness* de treinamento da população ($> 10^{-5}$), e 3. a queda do *fitness* de validação da população (5%). Os experimentos foram realizados com a utilização de quatro séries temporais: uma

artificial — Série *Random Walk*; e três reais — Série das Manchas Solares (*Sunspot*), Série da Medida de Brilho de um Estrela e Série do Índice *Dow Jones Industrial Average* (DJIA).

A. Random Walk

Um modelo *Random Walk* é um modelo AR(1) não estacionário [1], [9] com coeficiente auto-regressivo igual a 1, dado por $Z_t = Z_{t-1} + a_t$, onde o a_t é um termo de ruído.

A partir deste modelo foi gerado um conjunto de 1000 pontos, onde o valor inicial para a série foi $Z_0 = 100$ e o termo aleatório a_t foi construído a partir de uma distribuição uniforme de números aleatórios no intervalo de $[-10,10]$.

O comportamento médio de previsão para a série de *Random Walk* exibido na Tabela I.

TABELA I
ERROS MÉDIOS DE PREVISÃO DA SÉRIE *Random Walk*.

Medidas	Rede 1		Rede 2		Rede 3	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
MSE	0.0020	0.0003	0.0020	0.0006	0.0019	0.0004
MAPE	8.5527	0.5732	8.8004	0.9836	8.5738	0.7256
U de Theil	1.5040	0.2173	1.5538	0.4331	1.4075	0.3203
POCID	52.8340	1.2838	52.2267	2.5677	51.8623	1.3556
ARV	0.0761	0.0162	0.0807	0.0235	0.0128	0.0021
AIC	-1364.87	44.95	-1370.19	69.20	-1437.68	111.32
BIC	-969.40	72.74	-992.75	91.85	-1163.06	250.46

Destas dez inicializações é possível se eleger o indivíduo mais apto segundo as medidas de desempenho. Tal indivíduo é uma RNA do modelo REDE 3 de arquitetura 9-6-1, e os valores alcançados para as respectivas medidas de desempenho são expostos da Tabela II. Um gráfico comparativo entre os dados reais (linha sólida) e a previsão (linha tracejada) para o conjunto de teste é mostrado na Figura 3.

TABELA II
ERROS DE PREVISÃO DA SÉRIE *Random Walk* — INDIVÍDUO MAIS APTO.

Arquitetura	MSE	MAPE	U de Theil	POCID	ARV	AIC	BIC
9-6-1	0.0014	7.8359	1.0884	50.6072	0.0122	-1537.97	-1357.59

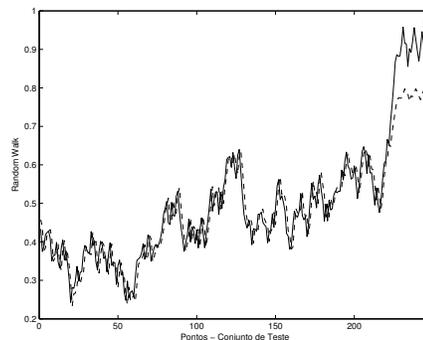


Figura 3. Gráfico de Previsão para a Série *Random Walk*.

B. Série das Manchas Solares

A série das manchas solares, ou *sunspots*, utilizada aqui são as totalizações anuais das manchas solares entre os anos de 1700 e 1988, formando um conjunto de 289 pontos.

Dez inicializações distintas com o sistema híbrido exposto levaram ao comportamento médio das medidas de desempenho para a previsão da série das Manchas Solares, como exposto na Tabela III.

TABELA III

ERROS MÉDIOS DE PREVISÃO DA SÉRIE DAS MANCHAS SOLARES.

Medidas	Rede 1		Rede 2		Rede 3	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
MSE	0.0070	0.0010	0.0073	0.0009	0.0071	0.0010
MAPE	107.8378	15.6144	109.5248	12.6494	111.1818	10.6874
U de Theil	0.6127	0.0899	0.6416	0.0823	0.6130	0.0900
POCID	51.1594	3.2800	51.7391	3.6184	52.8986	3.5003
ARV	0.1702	0.0287	0.1760	0.0221	0.0663	0.0133
AIC	-199.34	32.52	-185.05	21.83	-224.85	46.92
BIC	33.19	81.81	65.27	52.72	-34.69	125.04

O melhor indivíduo gerado a partir de todos os experimentos elegeu a arquitetura 10-4-1 para a RNA de modelo REDE 1, onde seus resultados para todas as medidas de desempenho são mostrados na Tabela IV. Um gráfico comparativo entre os dados reais (linha sólida) e a previsão (linha tracejada) para o conjunto de teste do melhor indivíduo é mostrado na Figura 4.

TABELA IV

ERROS DE PREVISÃO DA SÉRIE DAS MANCHAS SOLARES — INDIVÍDUO

MAIS APTO.

Arquitetura	MSE	MAPE	U de Theil	POCID	ARV	AIC	BIC
10-4-1	0.0063	106.0857	0.5549	50.7246	0.1351	-227.09	-29.8120

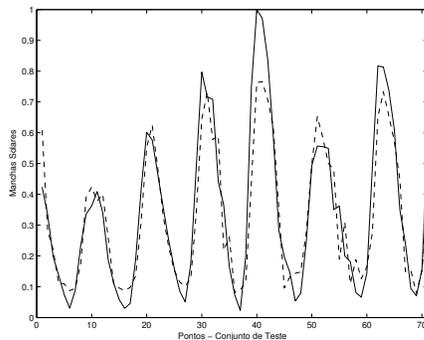


Figura 4. Gráfico de Previsão para a Série das Manchas Solares.

C. Série de Medidas do Brilho de uma Estrela

A série de medidas de brilhos de uma estrela é uma série natural com observações diárias, realizadas sempre no mesmo local e mesma hora, de uma estrela de brilho oscilante sendo constituída por 600 pontos. Esta série pode ser encontrada em <http://www-personal.buseco.monash.edu.au/~hyndman/TSDL>.

Dez experimentações distintas com o sistema híbrido exposto levaram ao comportamento médio das medidas de desempenho para a previsão da série de

medidas de brilhos de uma estrela, como exposto na Tabela V.

TABELA V

ERROS MÉDIOS DE PREVISÃO DA SÉRIE DO BRILHO DE UMA ESTRELA.

Medidas	Rede 1		Rede 2		Rede 3	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
MSE	0.0010	0.0003	0.0010	0.0003	0.0018	0.0002
MAPE	10.5140	2.8811	10.0609	2.2261	16.1847	3.1998
U de Theil	0.2604	0.0788	0.2541	0.0836	0.4728	0.0623
POCID	51.4286	1.2084	51.9728	1.0242	51.0884	2.2090
ARV	0.0138	0.0042	0.0136	0.0044	0.0057	0.0008
AIC	-868.49	47.63	-885.91	63.51	-790.83	32.28
BIC	-558.43	70.52	-600.60	101.54	-510.70	104.00

Destas experimentações é possível se escolher a de melhor desempenho segundo o conjunto de medias dos erros de previsão para o conjunto de teste. O indivíduo mais apto alcançou a arquitetura 10-7-1 para a RNA de modelo REDE 2, sendo os resultados expostos na Tabela VI. Um gráfico comparativo entre os dados reais (linha sólida) e a previsão (linha tracejada) para o conjunto de teste é mostrado na Figura 5.

TABELA VI

ERROS DE PREVISÃO DA SÉRIE DO BRILHO DE UMA ESTRELA — IND.

MAIS APTO.

Arquitetura	MSE	MAPE	U de Theil	POCID	ARV	AIC	BIC
10-7-1	0.0005	6.9045	0.1386	52.3810	0.0073	-983.70	-728.32

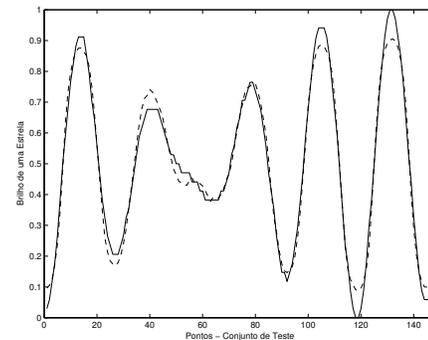


Figura 5. Gráfico de Previsão para a Série de Medidas do Brilho de uma Estrela.

D. Séries do Índice Dow Jones Industrial Average (DJIA)

A série do índice DJIA aqui utilizada é constituída por observações diárias do índice a partir de primeiro de janeiro de 1998 até 26 de agosto de 2003, totalizando 1420 pontos.

As dez inicializações distintas da metodologia híbrida resultou em um comportamento médio de previsão exposto na Tabela VII.

Dentre os experimentos, o indivíduo mais apto alcançou uma arquitetura 10-6-1 para a RNA de modelo REDE 2, com os seus resultados para as sete medidas de desempenho apresentados na Tabela VIII. Um gráfico comparativo entre os dados reais (linha sólida) e a previsão (linha tracejada) para o conjunto de teste é mostrado na Figura 6.

TABELA VII
VALORES DE TESTE DE PREVISÃO DA SÉRIE DO ÍNDICE *Dow Jones Industrial Average*.

Medidas	Rede 1		Rede 2		Rede 3	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
MSE	0.0018	0.0005	0.0018	0.0005	0.0022	0.0006
MAPE	5.4240	0.7212	5.3045	0.8001	5.9460	0.9127
U de Theil	2.1707	0.5426	2.0848	0.5700	2.5657	0.7403
POCID	51.6193	2.2571	49.4318	2.6311	50.7670	1.9593
ARV	0.0837	0.0222	0.0748	0.0159	0.0186	0.0040
AIC	-2053.91	96.31	-2071.67	109.92	-1965.80	101.72
BIC	-1633.69	114.50	-1654.86	156.36	-1465.81	101.07

TABELA VIII
VALORES DE TESTE DA SÉRIE DOW JONES PARA O INDIVÍDUO MAIS APTO.

Arquitetura	MSE	MAPE	U de Theil	POCID	ARV	AIC	BIC
10-6-1	0.0011	4.1980	1.2970	50.0000	0.0453	-2272.89	-1966.48

V. CONCLUSÕES

Neste trabalho foi exposto uma metodologia híbrida com a combinação de AG modificado com RNA's aplicada ao problema de previsão de séries temporais, sendo o AG utilizado inspirado no trabalho de Leung et al [4], que comprovou uma convergência mais rápida do método de busca com a utilização dos operadores de cruzamento e mutação expostos na Seção III.

Um ponto relevante deste trabalho é a utilização de um conjunto de sete medidas de desempenho para a análise dos resultados de previsão aplicados a quatro séries temporais complexas, tanto artificiais como do mundo real, bem como a utilização de três formas distintas para a modelagem das RNA's.

Analisando os resultados obtidos nos experimentos é possível notar que os erros *MSE* são relativamente baixos, uma vez que devido a normalização dos dados o maior valor possível é 1.0 (ou 100%), e de forma geral, foram encontrados valores da ordem de 10^{-3} (ou 0.1 %).

Entretanto, o comportamento da previsão normalizado quanto ao valor real da série, o erro *MAPE*, é um melhor parâmetro para construir um sentimento do desvio obtido pelo método. De forma geral, foram alcançados resultados para o *MAPE* da ordem de 10%, que na prática pode vir a gerar uma certeza de 90% a 80% na previsão.

Um fato interessante vem das medidas da estatística U de Theil e *POCID*. Bons resultados foram alcançados para as séries de fenômenos naturais e fracos resultados para as séries de *Random Walk* e Índice DJIA, ou seja, séries do tipo financeiras, visto que o modelo de *Random Walk* é adotado como um modelo financeiro simplista [9], mostrando indícios de comportamentos distintos para estas duas famílias de séries. Embora haja este comportamento distintos para as séries, a media ARV apresentou resultados sempre abaixo de 1.0, mostrando que os modelos gerados são superiores à simples previsão da média da série.

Já para os critérios de informação adotados, *AIC* e *BIC*, servem basicamente para a comparação entre modelos de uma mesma série temporal, ficando sem sentido a comparação direta dos mesmos para séries diferentes.

Desta forma, a metodologia deste trabalho demons-

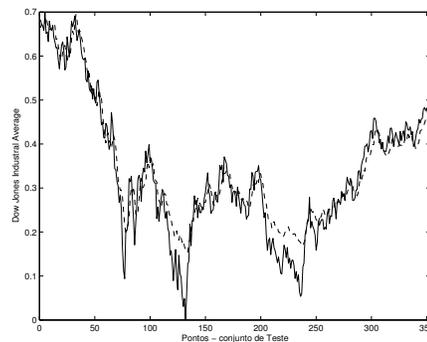


Figura 6. Gráfico de Previsão para a Série do Índice DJIA.

trar ser um opção válida para a abordagem ao problema de previsão de séries temporais com a utilização de um modelo não-linear, a RNA, detendo um custo de implementação bem inferior quando comparado ao custo de implementação de algoritmos convencionais de treinamento das RNA, como por exemplo o *Backpropagation* [10], além do fato do ajuste automático dos parâmetros da RNA.

Quanto a escolha do modelo para a RNA, os experimentos mostram que as três abordagens são bem semelhantes, contudo é possível se eleger o modelo REDE 2 como mais adequado de forma geral, uma vez que este foi escolhido em 50% dos experimentos e é a modelagem de menor custo computacional.

Entretanto, mais testes estão sendo realizados com esta metodologia com o uso de outras séries temporais, tanto artificiais como reais, em particular séries econômico-financeiras, para uma melhor caracterização dos custos computacionais envolvidos e a eficiência efetiva do método para aplicações práticas de previsão e modelagem de sistemas de apoio a decisão.

REFERÊNCIAS

- [1] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel, *Time Series Analysis: Forecasting and Control*, Prentice Hall, New Jersey, third edition, 1994.
- [2] Simon Hakykin, *Redes Neurais - Princípios e Prática*, Bookman, Porto Alegre - Brasil, 2a edition, 2002.
- [3] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, pp. 35–62, 1998.
- [4] F. H. F. Leung, H. K. Lam, S. H. Ling, and P. K. S. Tam, "Tuning of the structure and parameters of the neural network using an improved genetic algorithm," *IEEE Transaction on Neural Networks*, vol. 14, no. 1, pp. 79–88, January 2003.
- [5] R. Reed, "Pruning algorithms - a survey," *IEEE Transaction on Neural Networks*, vol. 4, no. 5, pp. 740–747, 1993.
- [6] Floris Takens, "Detecting strange attractor in turbulence," in *Dynamical Systems and Turbulence*, A. Dold and B. Eckmann, Eds., New York, 1980, vol. 898 of *Lecture Notes in Mathematics*, pp. 366–381, Springer-Verlag.
- [7] M. P. Clements and D. F. Hendry, "On the limitations of comparing mean square forecast errors," *Journal of Forecasting*, vol. 12, no. 8, pp. 617–637, Dec. 1993.
- [8] Chris Chatfield, *Time Series Forecasting*, Chapman & Hall/CRC, London, 2001.
- [9] Terence C. Mills, *The Econometrics Modelling of Financial Time Series*, Cambridge University Press, Cambridge, 2nd. edition, 2003.
- [10] F. S. Wong, "Time series forecasting using backpropagation neural networks," *Neurocomputing*, vol. 2, pp. 147–159, 1991.