

Previsão de séries de tempo: redes neurais artificiais e modelos estruturais*

LUIZ GUSTAVO L. FERNANDES **

PHILIPPE O. A. NAVAUX **

MARCELO SAVINO PORTUGAL ***

Neste artigo procura-se atingir um duplo objetivo: por um lado, o modelo de Redes Neurais Artificiais (RNA) é apresentado dando destaque às suas características estatísticas; e, por outro, são apresentados dois exercícios de previsão empíricos onde se tenta comparar o desempenho das RNAs e do modelo de séries de tempo estrutural, que pertence à classe dos modelos de decomposição em componentes não-observáveis (UCM). Na verdade, os modelos econométricos tradicionais são usados para fornecer uma base inicial para a determinação da arquitetura da RNA. Os resultados mostram que, para as previsões um passo à frente, não é possível discriminar inequivocamente a favor de um dos modelos; contudo, para previsões com horizontes de tempo mais longos, a RNA apresenta maior capacidade preditiva.

1 - Introdução

As Redes Neurais Artificiais (RNA) funcionam conceitualmente de forma similar ao cérebro humano, tentando reconhecer regularidades e padrões de dados, e são capazes de aprender com a experiência e fazer generalizações baseadas no seu conhecimento previamente acumulado. Embora biologicamente inspiradas, elas encontraram aplicações em diferentes áreas científicas. Neste trabalho, exploramos a habilidade das RNAs em gerar previsões para séries de tempo.

Em primeiro lugar, explora-se a ligação entre as RNAs e os modelos econométricos tradicionais e procura-se mostrar que elas são basicamente equivalentes ao problema de ajustamento de uma função não-linear a um conjunto de dados, onde a forma desta função é dada pela arquitetura da rede; e, em segundo, discute-se uma metodologia que facilite a determinação da arquitetura ou topologia da rede neural a ser utilizada, baseada nos resultados gerados pelos modelos econométricos tradicionais, tais como Arima e estrutural.¹

* Trabalho apresentado no XVII Encontro Brasileiro de Econometria, realizado em Salvador, 1995. Os autores gostariam de agradecer o auxílio dos bolsistas Frederico Augusto Pinto (CNPq/UFRGS) e Leandro Militsky (Pruni/UFRGS), bem como o apoio financeiro do CNPq (301445/92-3).

** Do Instituto de Informática da Universidade Federal do Rio Grande do Sul.

*** Do Departamento de Economia da Universidade Federal do Rio Grande do Sul.

¹ Para uma introdução às RNAs, ver, por exemplo, Portugal (1995) e Fernandes e Portugal (1995).

Além desta Introdução, o trabalho apresenta mais quatro seções: a Seção 2 tece considerações sobre a associação entre RNAs e os métodos estatísticos; a Seção 3 aborda a proposta de metodologia deste trabalho; a Seção 4 apresenta dois exemplos empíricos, onde é utilizada a metodologia proposta e são geradas previsões para séries de tempo, apresentando-se, ainda, uma comparação da *performance* preditiva das RNAs e dos modelos estruturais de séries de tempo; e, finalmente, a Seção 5 apresenta as considerações finais a respeito do trabalho desenvolvido e algumas sugestões de futuros tópicos de pesquisa.

2 - RNAs e estatística

Ao contrário do que possa parecer à primeira vista, as RNAs têm vários pontos de contato com os modelos estatísticos e econométricos tradicionais. Estas ligações acabam embotadas pelo uso de jargões técnicos distintos. Na Tabela 1, tentamos fornecer um dicionário reduzido com o objetivo de facilitar a comunicação entre economistas/econometristas e conexionistas. Vamos ainda tentar definir as RNAs utilizando-nos de uma linguagem mais usual do ponto de vista estatístico. Uma RNA pode ser definida como uma forma de mapear um número de entradas $(x_1, x_2, x_3, \dots, x_r)$ em um grupo de saídas $(0_1, 0_2, 0_3, \dots, 0_p)$. Para simplificar, imaginemos que $p = 1$, isto é, existe apenas um neurônio na camada de saída.

Em cada neurônio j da camada oculta, a entrada corresponde a uma soma ponderada representada por:

$$\sum_{i=0}^r y_{ji} x_i = y_{j0} + y_{j1} x_1 + y_{j2} x_2 + \dots + y_{jr} x_r$$

TABELA 1

Dicionário de RNAs — Econometria

RNAs	Estatística
Pesos	Parâmetros
Conjuntos de treinamento	Amostra
Entradas	Variáveis exógenas
Saídas	Variáveis endógenas
Retropropagação	Aproximação estocástica
Treinamento ou aprendizado	Estimação
Sinal de entrada	Valor das variáveis exógenas
Sinal de saída	Valor estimado
Alvo	Valor da variável endógena

O primeiro elemento da soma pode ser visto como o “viés”, que aparece em função da existência de um neurônio ligado a todos os demais da rede cujo sinal de saída é sempre igual à unidade ($x_0 = 1$). Assim, após a aplicação da função de ativação G , que pode ser uma função de distribuição acumulada como a logística, temos a saída de cada neurônio j da camada oculta dada por:

$$h_j = G \left(\sum_{i=0}^r y_{ji} x_i \right)$$

Assumindo a existência de q neurônios na camada oculta e fazendo $h_0 = 1$, temos então a entrada do neurônio da camada de saída representada por:

$$\sum_{j=0}^q \beta_j h_j = \beta_0 + \beta_1 h_1 + \beta_2 h_2 + \dots + \beta_q h_q$$

Neste sentido, o sinal de saída da RNA pode ser obtido como:

$$o = G \left(\sum_{j=0}^q \beta_j h_j \right)$$

ou, simplesmente:

$$o = G \left[\sum_{j=0}^q \beta_j \cdot G \left(\sum_{i=0}^r y_{ji} x_i \right) \right] = f(x, \theta) \quad (1)$$

No Gráfico 1 temos a representação de uma RNA com quatro neurônios na camada de entrada, onde o triângulo representa o “viés”, três neurônios na camada intermediária ou oculta e um neurônio na camada de saída. O vetor de pesos y estabelece as ligações entre as camadas de entrada e oculta, enquanto o vetor β estabelece as ligações entre as camadas oculta e de saída. Na equação (1) temos a saída da RNA como uma função das entradas da rede e dos diversos pesos. Podemos escrever resumidamente $f(x, \theta)$, onde x é o vetor de entradas, θ o vetor dos pesos y e β e f pode ser chamada de “função de saída da rede”, e sua forma precisa vai depender da arquitetura usada na rede. Neste sentido, podemos dizer que a procura por uma arquitetura ótima para uma RNA resume-se à procura da forma funcional f apropriada. Existem então duas questões a serem resolvidas: a escolha da forma funcional f , que está associada à arquitetura da rede, isto é, ao número de camadas da rede e ao número de neurônios em cada camada, e a estimação do vetor de parâmetros θ , que é feita pela utilização do método de retropropagação.²

² Detalhes sobre o método de retropropagação podem ser obtidos em Wasserman (1989).

Gráfico 1
Representação de uma RNA

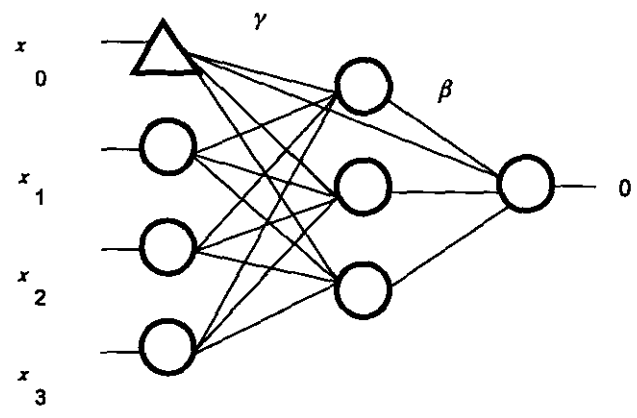
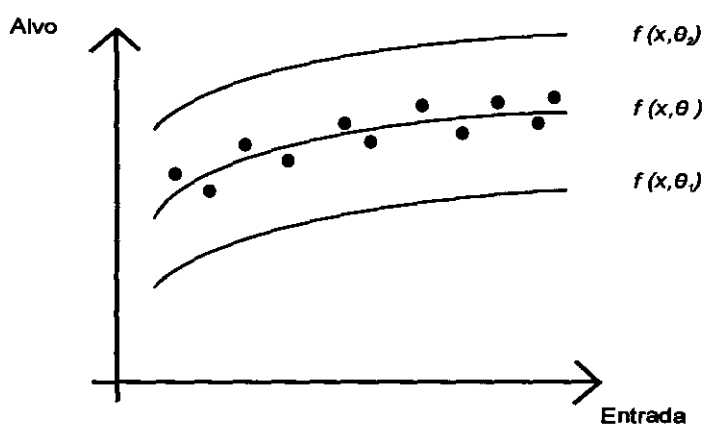


Gráfico 2
Ajuste de mínimos quadrados não-lineares



Consideremos agora o problema econométrico usual de se estimar uma regressão não-linear, em que a expressão $f(x, \theta)$ — onde x é o vetor de variáveis exógenas e θ o vetor de parâmetros — representa uma família de curvas cuja forma funcional é definida *a priori*, e o problema consiste em obter o estimador $\hat{\theta}$ ótimo, isto é, aquele que minimiza a soma dos quadros dos resíduos. Em outras palavras, o problema consiste em obter um membro específico desta família de curvas $f(x, \hat{\theta})$ que apresenta o melhor ajuste aos dados.

Em White (1992) é mostrado que o estimador de mínimos quadrados não-lineares ($\hat{\theta}$) tende assintoticamente para os pesos ótimos aprendidos pela RNA (θ^*). É neste sentido que se pode afirmar que a grande novidade introduzida pelas RNAs consiste em focalizar a atenção na forma funcional f . Elas têm-se mostrado extremamente úteis na resolução de problemas de reconhecimento de padrões, classificação, previsão, etc., pois possuem a propriedade de ser "aproximações universais". Vale dizer, uma RNA, com um número suficiente de neurônios na camada oculta e uma função de ativação tipo logística, consegue aproximar, com grau de precisão desejado, qualquer mapa de um espaço dimensional finito em um outro espaço dimensional finito qualquer.³ Já os procedimentos de aprendizagem utilizados para obter os pesos consistem em um processo de estimação usual em estatística. Nas palavras de White (1992, p. 87): "neural networks procedures are inherently statistical techniques". O algoritmo de retropropagação é utilizado simplesmente para determinar o mínimo local da superfície de erros.⁴

Robbins e Monro (1951) propuseram um procedimento para estimar o vetor $\hat{\theta}$ que resolve a equação $E(m(Z_t, \theta)) = 0$, onde E representa a esperança matemática, θ é um vetor de parâmetros e $m(Z_t, \theta)$ é uma quantidade aleatória, cuja aleatoriedade advém da presença da variável aleatória Z_t . A única informação conhecida para obter $\hat{\theta}$ é apenas uma realização da variável aleatória Z_t , para $t = 1, 2, \dots, n$. Robbins e Monro (1951) propuseram estimar θ de forma recursiva pela seguinte equação:

$$\theta_t = \theta_{t-1} + \eta_t \cdot m(Z_t, \theta_{t-1}) \quad \text{para } t = 1, 2, \dots, n$$

onde θ_0 é escolhido aleatoriamente. Fazendo-se:

$$\eta_t = \eta$$

$$m(Z_t, \theta_{t-1}) = \nabla f(x_t, \theta_{t-1}) \cdot (y_t - f(x_t, \theta_{t-1}))$$

$$Z_t = (x_t, y_t)$$

obtem-se:

$$\theta_t = \theta_{t-1} + \eta \cdot \nabla f(x_t, \theta_{t-1}) \cdot (y_t - f(x_t, \theta_{t-1}))$$

que é, na verdade, uma forma genérica de escrever o algoritmo de retropropagação.

³ A prova pode ser encontrada em White (1992, Cap. 3).

⁴ Deve ficar claro, a partir da apresentação das equações para o algoritmo de retropropagação, que estamos tratando com uma variação do método de Newton-Raphson.

3 - Discussão da metodologia para definição da arquitetura

Esta seção apresenta a metodologia proposta para a modelagem da arquitetura (topologia) da rede neural (Perceptron Multi-Camadas) que seja capaz de realizar prognósticos confiáveis sobre valores futuros de séries temporais. A questão a que se procurou responder foi: qual o tamanho ótimo para a rede de treinamento? Neste sentido, são apresentados resumidamente a seguir dois estudos empíricos que se dedicaram a pesquisar soluções para a questão da topologia da rede.

3.1 - Determinação da camada de entrada da rede

Em Varfis e Versino (1990), que tratam da previsão de séries econômicas mensais, encontra-se uma proposta de metodologia onde se procura modelar a entrada da rede de maneira a captar as componentes da série temporal. Desta forma, as unidades foram divididas em três grupos:

- a) valores passados da série;
- b) valores passados referentes ao mesmo mês em anos anteriores, visando captar as tendências ou a sazonalidade da série;
- c) grupo de 12 unidades binárias para captar a sazonalidade da série, tais como 100000000000 para janeiro, 010000000000 para fevereiro, e assim sucessivamente, até 000000000001 para dezembro.

3.2 - Determinação da camada escondida de rede

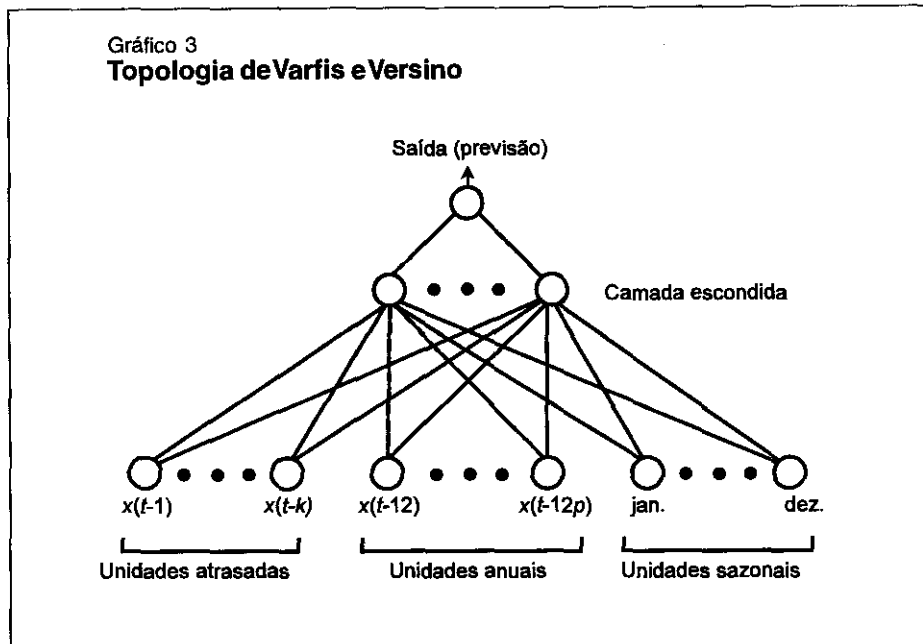
Muitos autores, como Caudill (1990), apresentam o número de unidades das camadas escondidas como incógnita e simplesmente afirmam que “o tamanho da camada escondida é escolhido usando o bom senso: se a camada for muito grande, a rede estará memorizando os padrões e com isso perdendo a capacidade de generalização; por outro lado, se a camada for muito pequena, a rede levará muito mais interações para atingir a precisão desejada”. Note-se que a memorização no contexto de um modelo de RNAs é equivalente ao problema econométrico da superparametrização do modelo. Quando a RNA memoriza um certo conjunto de dados, ela vai apresentar pequenos erros de previsão para o período amostral, mas grandes erros para previsões fora da amostra.

Já Rumelhart e Weigend (1990) afirmam que “a rede que melhor generaliza os dados é a menor rede apta a realizar o treinamento dos dados”. Neste sentido, os autores propõem um método que é uma extensão do método do gradiente descendente, onde a idéia básica é começar com uma rede muito grande, mas associando um custo a cada conexão da rede, o que seria decisivo para a eliminação ou não de um determinado peso ou conexão.

A função custo proposta é a soma de dois termos: o primeiro é a soma dos quadrados dos erros enquanto o segundo descreve um custo para cada peso na rede. Se um certo

desempenho de um conjunto de dados para treinamento pode ser obtido com poucos pesos, a função custo irá encorajar a redução do número de conexões e, eventualmente, eliminar tantos pesos quantos forem possíveis.

O Gráfico 3 apresenta de forma esquemática a topologia de uma RNA utilizando a abordagem de Varfis e Versino (1990).



3.3 - Utilizando as características não-observáveis da série

Usando o modelo estrutural de séries de tempo [ver Harvey (1989)] ou o modelo linear dinâmico bayesiano [ver West e Harrison (1989)] é possível decompor uma série em seus componentes não-observáveis diretamente, tais como tendência, ciclo, sazonalidade, irregular etc. Uma vez conhecidas estas características básicas da série, é possível construir um ou mais neurônios que busquem representar estas características não-observáveis da série. Assim sendo, podemos usar a modelagem econométrica de componentes não-observáveis como uma fase preliminar na montagem da camada de entrada da RNA e utilizar a observação feita por Rumelhart e Weigend (1990) com o objetivo de estabelecer o número de neurônios da camada escondida da arquitetura. Vale dizer, a busca da melhor configuração para a camada escondida da rede é feita de forma a minimizar o número de neurônios nesta camada sem que haja perda de capacidade preditiva.

4 - Resultados

Serão apresentados a seguir dois experimentos de utilização de RNAs para previsão de séries temporais utilizando a metodologia de determinação da arquitetura (topologia) da rede proposta neste trabalho. O método estatístico utilizado para obter informações a respeito das componentes das séries também fornecerá, através dos resultados de suas previsões, um *benchmark* para que se possa avaliar a qualidade dos resultados obtidos pelas redes neurais.

Foram escolhidas para a realização do exercício empírico duas séries que apresentam características distintas e que já foram bastante estudadas na literatura de previsão. Desta forma, procuramos facilitar a comparação e a avaliação dos resultados, bem como aplicar a metodologia proposta a diferentes tipos de problemas. As séries escolhidas foram:

série A: índice mensal de passageiros de linhas aéreas nos Estados Unidos;

série B: índice pluviométrico anual da cidade de Fortaleza.

Estas séries terão suas características analisadas separadamente a seguir, sendo também apresentadas as previsões realizadas para seus valores futuros. Para tanto, será utilizado o modelo estrutural clássico para previsão de séries temporais (UCM).⁵

4.1 - Modelagem e resultados da série A

Esta é uma das séries clássicas utilizadas em testes de modelos de previsão de séries temporais [ver Box e Jenkins (1976) e Harvey (1989)] e corresponde ao índice mensal de passageiros nas linhas aéreas dos Estados Unidos entre janeiro de 1960 e dezembro de 1971. O Gráfico 4, que mostra os valores (em mil unidades) de toda a série, deixa claro que ela apresenta a tendência e a sazonalidade como características mais marcantes. Para esta série, o conjunto de observações de janeiro de 1960 a dezembro de 1970 será utilizado para a estimação do modelo (fase de aprendizado), perfazendo um total de 132 observações, enquanto o período correspondente a 1971 será utilizado para as previsões (validação do modelo). Seguindo-se Harvey (1989), o modelo estrutural foi estimado com uma tendência estocástica e um componente sazonal estocástico modelado de forma trigonométrica.

O fator sazonal mostra um período de pico anual principal no número de passageiros que viajaram de avião nos Estados Unidos durante julho e agosto (período de férias de verão) e um pico secundário em março e abril, assim como um período de redução sazonal no número de passageiros de avião entre setembro e fevereiro, em função do inverno. O Gráfico 5 mostra ainda que a sazonalidade da série não é constante ao longo do tempo, tendo aumentado seguidamente após 1965.

⁵ As previsões para o modelo de séries de tempo estrutural foram realizadas usando-se o programa STAMP. Já para o aprendizado e previsões com RNAs utilizou-se o programa Neural Works II.

Gráfico 4
Índice mensal de passageiros de linhas aéreas – 1960/71

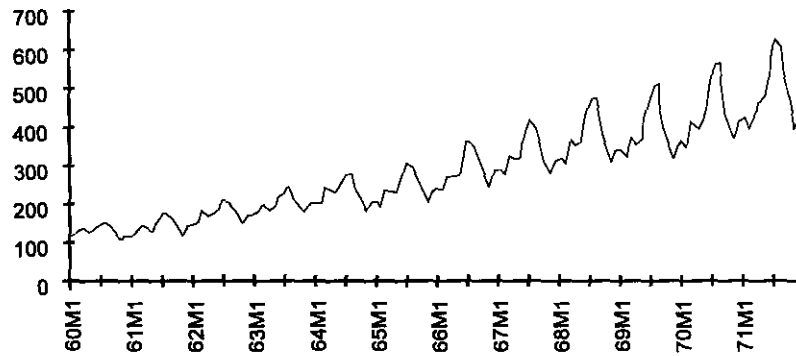
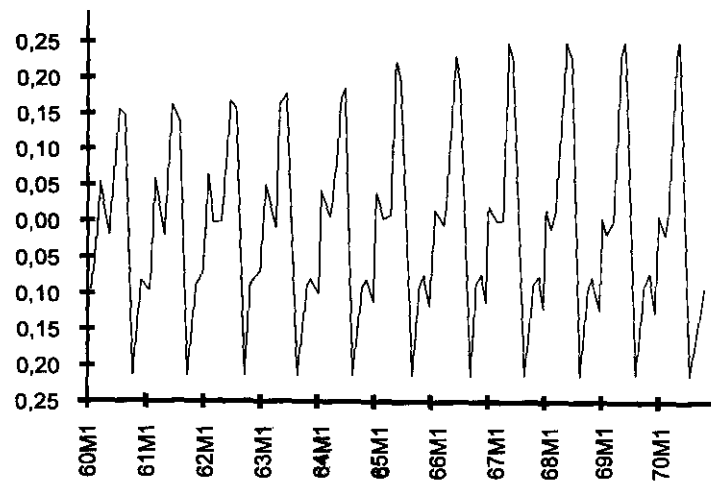
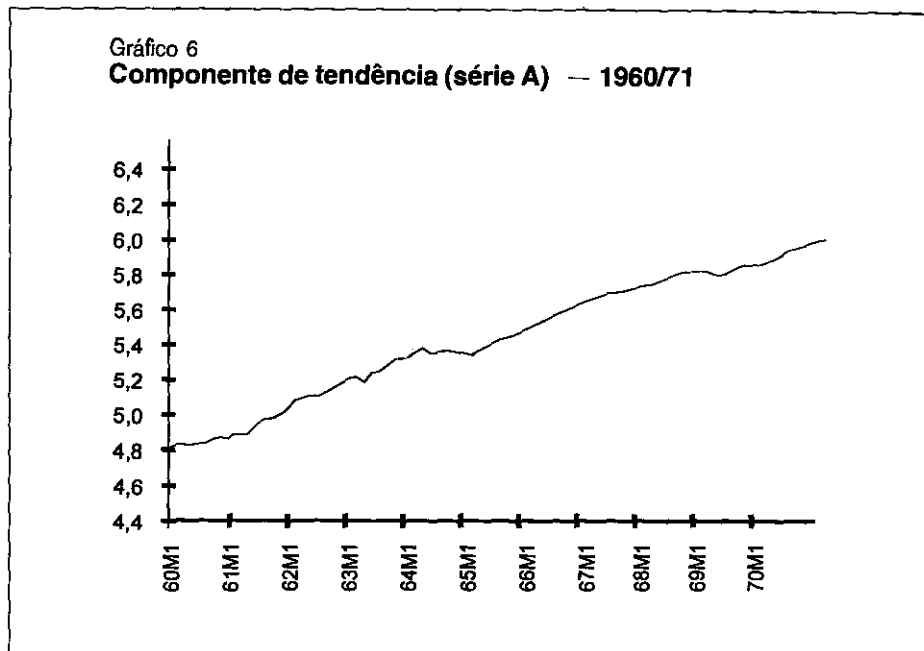


Gráfico 5
Componente sazonal (série A) – 1960/71





A componente tendência aponta para um crescimento linear da série, indicando, portanto, um crescimento médio da série praticamente constante ao longo do período estudado. De fato, ao analisarmos o Gráfico 4, podemos facilmente perceber o incremento linear no número de passageiros que viajam de avião a cada mês.

Por último, é apresentado no Gráfico 7 o componente irregular da série, que, em função do ótimo ajuste do modelo, possui valores sempre bastante próximos de zero. Chama a atenção, contudo, a redução da variância do componente irregular para o período após 1965.

Seguindo a proposta apresentada na seção anterior, a arquitetura que obteve os melhores resultados na realização dos prognósticos da série A é mostrada no Gráfico 8. Sua configuração apresenta 15 neurônios na camada de entrada, seis na intermediária e um na de saída. Segundo Zandonade (1993), que faz uma analogia com os modelos Arima, podemos denominar esta arquitetura RNA (15,6,1). A camada de entrada apresenta os seguintes neurônios:

- um neurônio contendo o valor da observação imediatamente anterior à atual que se quer prever (Y_{t-1});
- 12 neurônios binários (ver Subseção 3.1), utilizados com o objetivo de auxiliar a rede a identificar a componente sazonal da série A; e
- um neurônio contendo uma seqüência linearmente crescente, utilizado com o objetivo de auxiliar a rede a identificar a componente tendência da série (neurônio de tendência).

Gráfico 7
Componente irregular (série A) – 1960/71

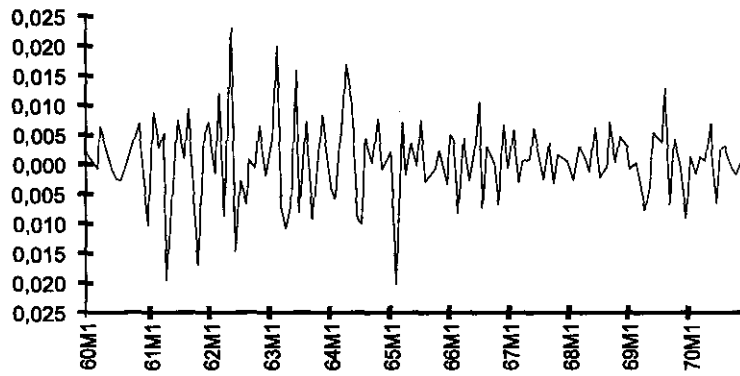
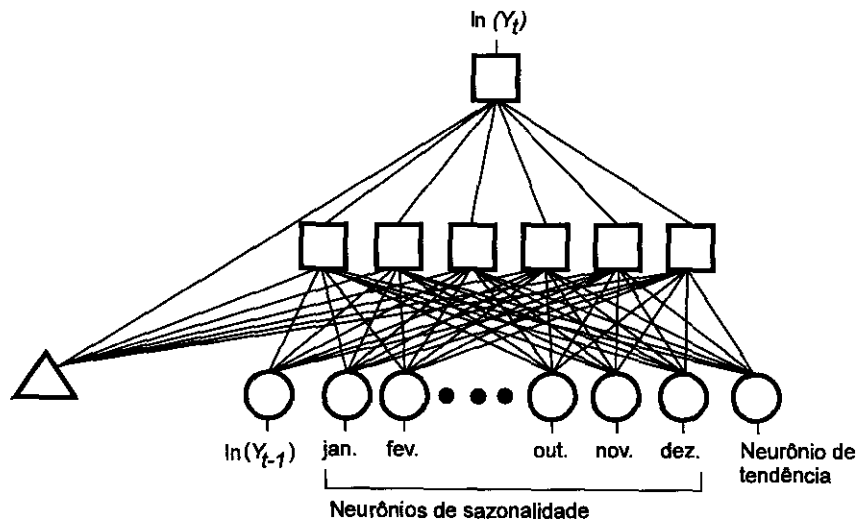


Gráfico 8
Arquitetura da RNA para a série A – $\ln(Y_t)$



Esta RNA foi então treinada a partir dos seguintes parâmetros: ⁶

- o conjunto de observações foi apresentado 15 mil vezes para a rede;
- constante de aprendizado $\alpha = 0,9$;
- termo de momento $\beta = 0,6$ nas 10 mil épocas iniciais e 0,9 nas 5 mil restantes; ⁷
- tamanho do intervalo de variação dos pesos igual a 2; e
- função de transferência do tipo sigmóide. ⁸

Os resultados obtidos para a previsão da série A podem ser visualizados na Tabela 2, onde se pode perceber que os erros apresentados nos prognósticos realizados apresentam seu ponto máximo em 10,58%, ficando, na sua grande maioria, abaixo de 5% tanto para a previsão um passo à frente como 12 passos à frente. ⁹

O fato de o modelo de RNAs ter conseguido resultados tão expressivos confirma sua capacidade de realizar prognósticos precisos. A Tabela 4 faz a comparação entre as previsões geradas pela RNA e pelo modelo estrutural (UCM). Apesar de os resultados obtidos pelas duas formas de previsão terem sido satisfatórios, convém ressaltar que a RNA está prevendo melhor para o horizonte mais longo, enquanto o modelo estrutural gera as melhores previsões um passo à frente.

Os Gráficos 9 e 10, que mostram as previsões geradas por ambos os modelos, permitem melhor visualização dos resultados, facilitando, portanto, a comparação. Um resultado semelhante a este, vale dizer, onde a RNA apresenta melhor desempenho para horizontes mais longos, foi também obtido por Portugal (1995) para a série de produção industrial do Estado do Rio Grande do Sul. ¹⁰

⁶ Para uma discussão mais detalhada dos parâmetros de treinamento, ver Portugal (1995) ou Fernandes (1995).

⁷ A utilização de termos de momento diferenciados ajuda a reduzir o tempo de treinamento.

⁸ Durante o treinamento, os dados de entrada e de saída (ênuplas de treinamento) são apresentados à rede de forma aleatória.

⁹ As previsões para mais de um passo à frente apresentadas neste artigo são, na verdade, uma sucessão de previsões com horizontes de previsão crescentes. Vale dizer, no caso das previsões 12 passos à frente apenas o último valor, referente a dezembro de 1971, é realmente uma previsão 12 passos à frente. Os demais são uma seqüência de previsões para horizontes de previsão crescentes.

¹⁰ Na verdade, este é um resultado muito encontrado na literatura, independentemente do modelo econométrico tradicional que se está utilizando para comparação. Swanson e White (1994), por exemplo, chegam também a esta conclusão comparando os resultados das RNAs com previsões geradas por modelos VAR.

TABELA 2

Resultados da RNA (série A) — 1971

Meses	Real	Um passo à frente	Erro percentual (%)	12 passos à frente	Erro percentual (%)
Janeiro	417	407,28	2,39	407,28	2,39
Fevereiro	391	368,12	6,22	377,89	3,47
Março	419	428,33	-2,19	439,57	-4,68
Abril	461	416,88	10,58	419,64	9,86
Maiο	472	455,27	3,67	460,31	2,54
Junho	535	522,33	2,43	536,73	-0,32
Julho	622	593,77	4,75	607,59	2,37
Agosto	606	589,87	2,73	626,34	-3,25
Setembro	508	499,50	1,70	499,65	1,67
Outubro	461	434,07	6,20	440,01	4,77
Novembro	390	396,99	-1,76	391,47	-0,38
Dezembro	432	422,38	2,28	424,79	1,70

TABELA 3

Resultados para o modelo estrutural (série A) — 1971

Meses	Real	Um passo à frente	Erro percentual (%)	12 passos à frente	Erro percentual (%)
Janeiro	417	423,86	-1,62	423,86	-1,62
Fevereiro	391	398,22	-1,81	401,79	2,69
Março	419	462,25	-9,36	471,26	-11,09
Abril	461	424,41	8,62	456,41	1,01
Maiο	472	463,17	1,91	477,09	-1,07
Junho	535	539,10	-0,76	546,15	-2,04
Julho	622	609,05	2,13	622,97	0,16
Agosto	606	628,41	-3,57	632,32	-4,16
Setembro	508	510,61	-0,51	525,95	-3,41
Outubro	461	449,93	2,46	463,31	-0,50
Novembro	390	402,82	-3,18	409,61	-4,79
Dezembro	432	422,26	-2,32	455,37	-5,13

Gráfico 9
Previsões um passo à frente

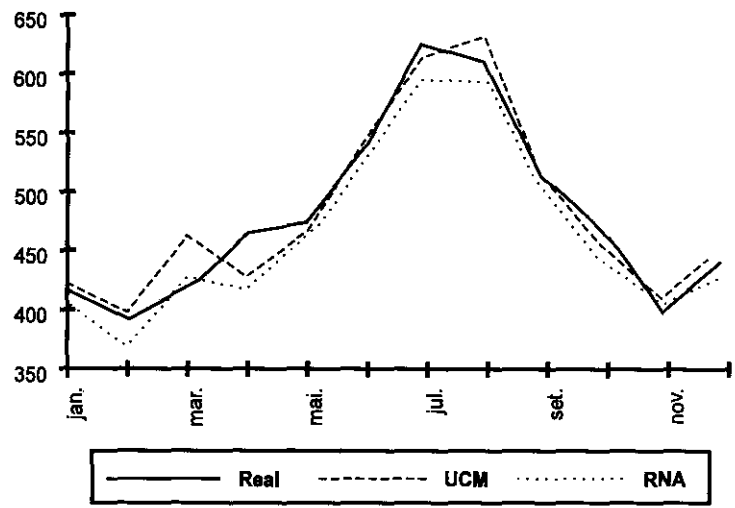


Gráfico 10
Previsões 12 passos à frente

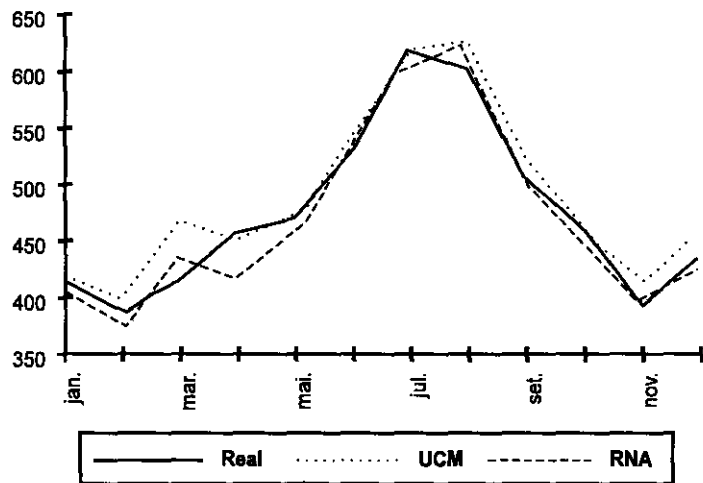


TABELA 4

Medidas de erro — RNA e UCM

	UCM — um passo à frente	UCM — 12 passos à frente	RNA — um passo à frente	RNA — 12 passos à frente
RMSE	4,19	4,26	4,94	3,39
MAE	3,19	3,14	3,91	3,12

4.2 - Modelagem e resultados da série B

Esta série apresenta um conjunto de 134 observações do índice pluviométrico anual, em centímetros, da cidade de Fortaleza, para o período 1849/1984. Como esta série é provavelmente a maior disponível do índice pluviométrico de uma região que sofre frequentemente com secas severas, ela tem sido objeto de um grande número de estudos na área de prognósticos de séries temporais (Gráfico 11) [ver Morettin, Mesquita e Rocha (1985) e Harvey (1989)].

Para esta série, o conjunto de observações entre 1849 e 1979 será utilizado como um ajuste do modelo (fase de aprendizado), ficando os cinco anos restantes (de 1980 a 1984) para a realização das previsões (validação do modelo). O Gráfico 11 deixa claro que estamos trabalhando com uma série estacionária cuja média é 142,2, além de apresentar também um comportamento cíclico. Na literatura existente já foi determinada a presença de pelo menos dois ciclos distintos, o primeiro correspondendo a um período de 13 anos e o segundo a um período de 26 anos.

Novamente seguindo Harvey (1989), o modelo estrutural foi estimado com um ciclo trigonométrico estocástico e uma constante para representar a média da série. Os resultados para as previsões um passo e cinco passos à frente encontram-se na Tabela 5, enquanto os componentes cíclico e irregular para a série B são apresentados nos Gráficos 12 e 13. Já uma inspeção dos Gráficos 16 e 17 pode levar o leitor a acreditar que um modelo com nível estocástico seria mais apropriado para os dados, na medida em que permitiria um ajuste do nível. Contudo, este modelo não apresenta nenhuma melhora significativa em relação àquele com nível fixo, tanto no que diz respeito à variância amostral como às previsões para o período 1980/84.

Gráfico 11
Índice pluviométrico de Fortaleza — 1849/1979

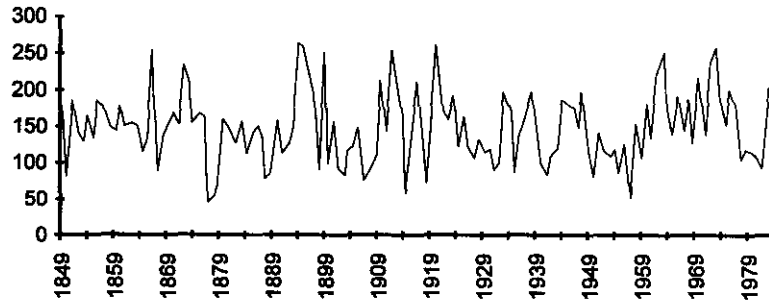


Gráfico 12
Componente cíclico (série B) — 1849/1979

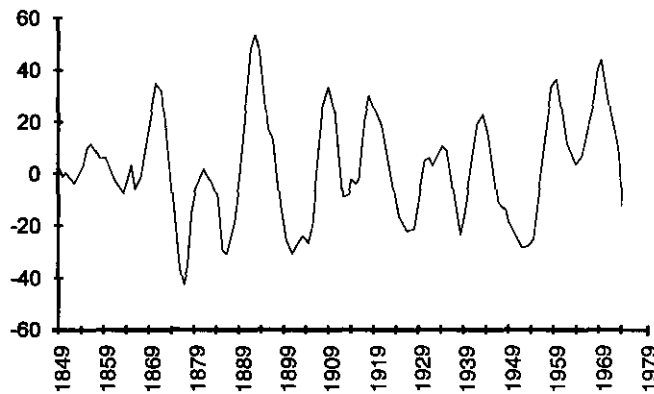


Gráfico 13
Componente irregular (série B) — 1849/1979

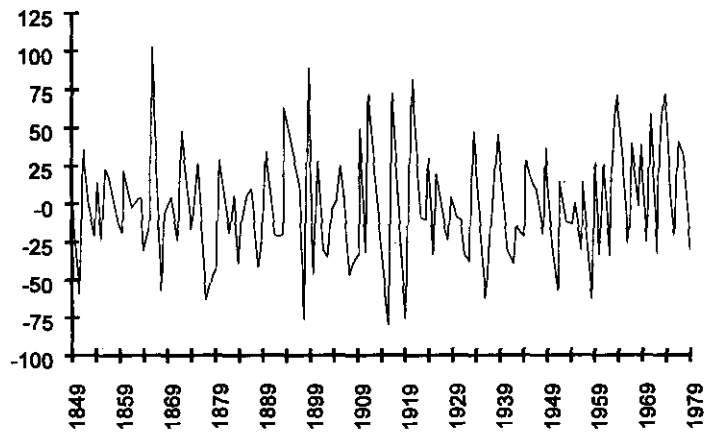


Gráfico 14
Neurônio cíclico usado na RNA — 1849/1979

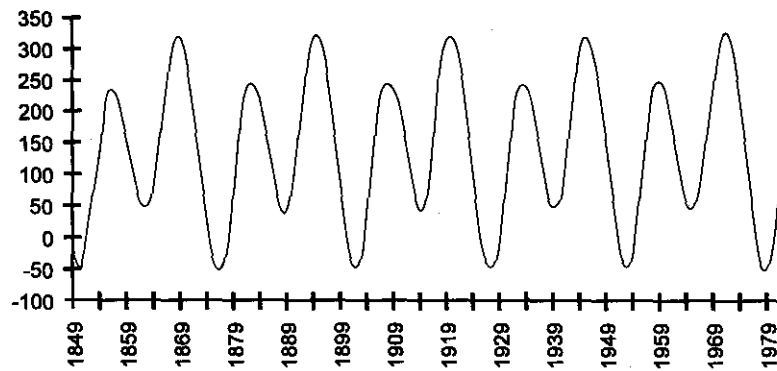


Gráfico 15
Arquitetura do modelo de RNA (série B)

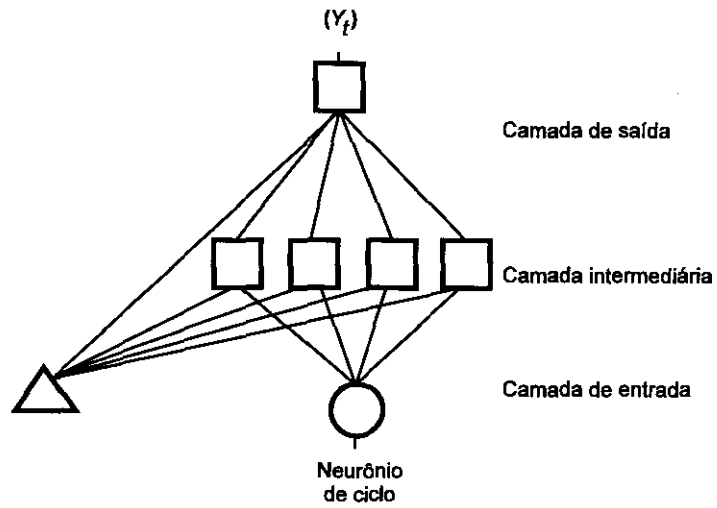
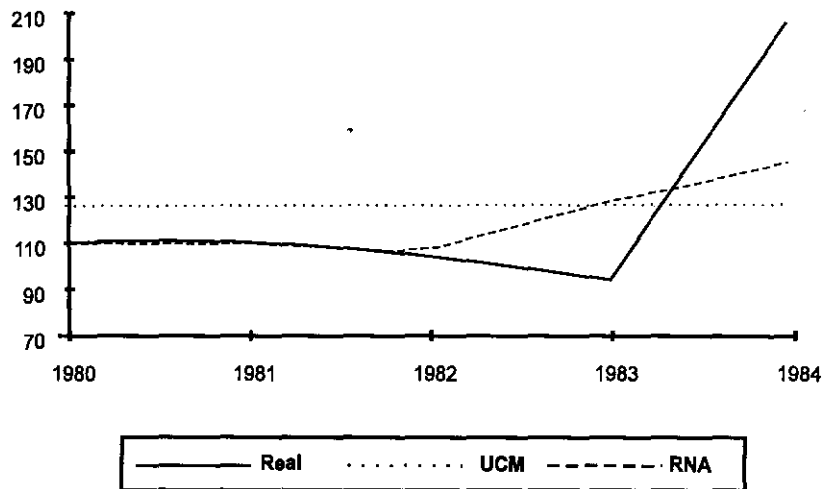


Gráfico 16
Previsões um passo à frente — 1980/84



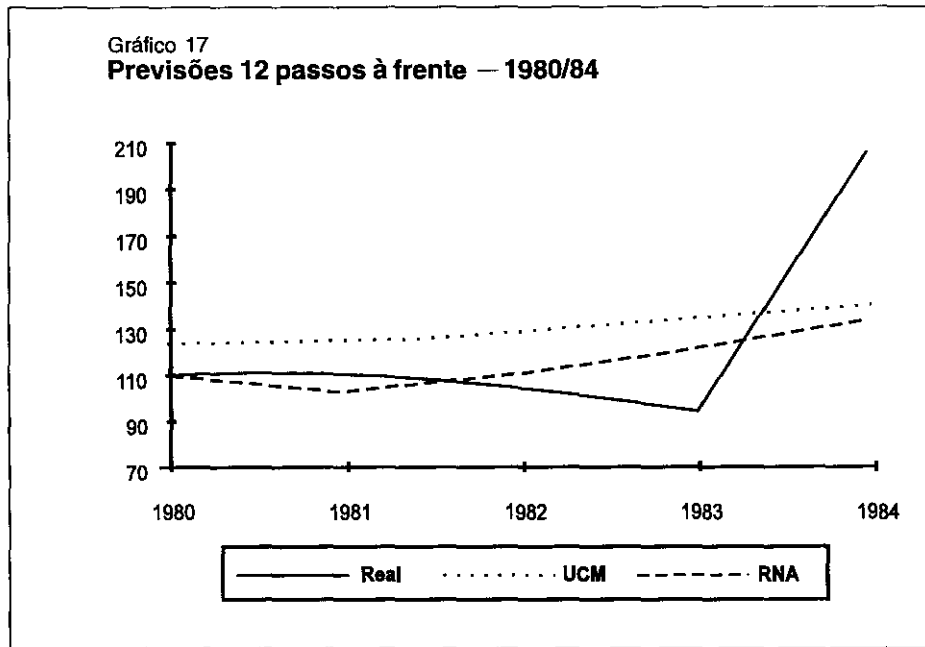


TABELA 5

Resultados do modelo estrutural (série B) — 1980/84

Período	Real	Um passo à frente	Erro percentual (%)	Cinco passos à frente	Erro percentual (%)
1980	109,5	127,47	-14,10	127,47	-14,10
1981	110,0	122,46	-10,17	126,36	-12,95
1982	100,4	122,74	-18,20	128,16	-21,66
1983	88,4	123,57	-28,46	131,74	-32,90
1984	198,1	123,67	37,57	136,00	31,35

A simples inspeção do Gráfico 13 indica a grande dificuldade enfrentada para a realização de prognósticos que apresentem pequenos erros. Note-se, por exemplo, que em algumas observações o valor do ruído aleatório pode atingir até 50% do índice.

Conforme proposto na seção anterior, utilizaremos os resultados do modelo estrutural para auxiliar na montagem da primeira camada da RNA. Em função da estacionaridade da série e da presença de um componente cíclico, usamos apenas dois neurônios na camada de entrada da rede, um para representar a constante e outro para representar o ciclo. Os valores para as entradas do neurônio cíclico foram obtidos a partir do ciclo estimado por Morettin, Mesquita e Rocha (1985) apresentado na equação (2):

$$Y_t = 142 + 20,3 \cdot \cos(0,24_t + 1,99) + 25,6 \cdot \cos(0,48_t + 2,2)$$

onde Y_t assume os valores do ciclo conforme a variação de t no tempo. Utilizando a equação (2), podemos gerar os dados para o ciclo com duas ondas apresentado no Gráfico 14.

O Gráfico 15 mostra a arquitetura escolhida da rede. Note-se que foram necessários quatro neurônios na camada intermediária para que a rede não perdesse sua capacidade de generalização. A configuração final da arquitetura é, portanto, RNA (2,4,1).

A arquitetura do Gráfico 15 foi treinada a partir dos seguintes parâmetros:

- o conjunto de observações foi apresentado 20 mil vezes para a rede;
- constante de $\alpha = 0,9$;
- termo de momento $\beta = 0,6$ nas 15 mil épocas iniciais e 0,9 nas 5 mil restantes;
- tamanho do intervalo de variação dos pesos igual a 2; e
- função de transferência do tipo sigmóide.

Ao contrário do que seria esperado, em função da análise realizada do componente irregular da série gerado pela estimação do modelo estrutural, os resultados das previsões utilizando rede neural para a série B, mostrados na Tabela 6, são muito satisfatórios, apesar dos erros bastante significativos para 1983 e 1984, anos que foram de difícil previsão, pois houve uma grande seca em 1983 seguida de um ano com uma grande quantidade de chuvas. Em 1983, o índice pluviométrico foi 37,8%, abaixo da sua média histórica, enquanto em 1984 as chuvas ficaram 39,5% acima da média. Temos, portanto, uma variação no índice pluviométrico entre 1984 e 1985 de 124,1%.

Contudo, para os três primeiros anos do período de previsão, tanto um passo à frente quanto cinco passos à frente, a RNA apresentou resultados bastante satisfatórios, com erros percentuais abaixo de 4%, indicando sua capacidade de aprender o ciclo presente nos dados.

TABELA 6

Resultados das RNAs (série B) — 1980/84

Período	Real	Um passo à frente	Erro percentual (%)	Cinco passos à frente	Erro percentual (%)
1980	109,5	109,67	-0,16	109,67	-0,16
1981	110,0	108,92	-0,98	103,94	5,51
1982	100,4	104,35	-3,78	107,63	-6,72
1983	88,4	122,34	-27,74	118,97	-25,69
1984	198,1	137,40	30,64	132,65	33,04

A Tabela 7, onde são apresentados os resultados comparativos entre os dois modelos, deixa claro que, para a série B, a capacidade preditiva da rede neural superou a do modelo estrutural, tanto para horizontes curtos como longos de tempo. Mesmo que ambos tenham tido erros bastante consideráveis para o final do período, a rede apresentou três prognósticos com uma precisão razoável, enquanto o UCM em momento algum mostrou erro abaixo de 10%.

TABELA 7

Medidas de erro RNA x UCM

	UCM— um passo à frente	UCM— cinco passos à frente	RNA— um passo à frente	RNA— cinco passos à frente
RMSE	23,90	24,09	18,57	19,12
MAE	21,70	22,59	12,66	14,22

5 - Considerações finais e pesquisas futuras

Este estudo não teve a pretensão de estabelecer que as RNAs, mais especificamente o modelo Perceptron Multi-Camadas com aprendizado por retropropagação de erros, apresentam-se como o melhor método de realização de previsão de valores futuros de séries temporais, uma vez que o trabalho empírico não é abrangente e os modelos não são comparados teoricamente. Porém, em face dos resultados apresentados, pode-se dizer com segurança que as RNAs são uma ferramenta poderosa para a realização de previsão de séries temporais, capazes de realizar prognósticos com o mesmo nível de precisão ou, em alguns casos, até maior que o modelo de séries de tempo estrutural. Na verdade, este artigo soma-se a uma

literatura empírica que já está se tornando volumosa, onde as RNAs aparecem como o melhor método de previsão para horizontes temporais mais longos. O caminho óbvio, no sentido de realizar uma comparação mais extensiva entre os dois modelos, envolveria a realização de experimentos de Monte Carlo com estimação de “superfícies de resposta” [ver Hendry (1984)].

A principal dificuldade na utilização de RNAs na previsão de séries temporais ainda é a determinação de sua arquitetura ótima. Ainda não se possui uma metodologia consistente que apresente a melhor configuração da rede para cada série proposta. A metodologia proposta por este trabalho avança no que diz respeito à determinação da camada de entrada da rede, mas apresenta como seu fator limitante a determinação do número de neurônios da camada intermediária. Uma sugestão que pode vir a ser estudada em trabalhos futuros é a utilização de modelos de RNAs com poda automática de neurônios que apresentem menor importância ao longo do aprendizado da rede, despertando particular interesse neste sentido o Modelo Neural Combinatório [ver Machado e Rocha (1992)].

Abstract

This article has a dual objective: on one hand we present the Artificial Neural Network (ANN) model from a statistical perspective; on the other hand, we also present two forecasting exercises to allow an empirical comparison between the ANN model and the traditional structural time series or unobservable components model (UCM). We use the UCM as a way to help determining the topology of the ANN. The results do not show clear cut conclusion as far as one step ahead forecasts are concerned, but for longer time horizons the ANN has a better forecasting performance.

Bibliografia

BOX, G. E. P., JENKINS, G. M. *Time series analysis, forecasting and control*. San Francisco: Holden-Day, 1976.

CAUDILL, M. *Neural network primer*. Miller Freeman Publications, 1990.

FERNANDES, L. G. *Utilização de redes neurais na análise e previsão de séries temporais*. Porto Alegre: Instituto de Informática/UFRGS, 1995 (Dissertação de Mestrado).

FERNANDES, L. G., PORTUGAL, M. S. *Redes neurais artificiais e previsão de séries econômicas: uma introdução*. Porto Alegre: Curso de Pós-Graduação em Economia/UFRGS, 1995 (Textos para Discussão, 95/1).

- HARVEY, A. C. *Forecasting structural time series models and the kalman filter*. Cambridge: Cambridge University Press, 1989.
- HENDRY, D. F. Monte Carlo experiments in econometrics. In: GRILICHES, Z., INTRILIGATOR, M. D. *Handbook of Econometrics*, 1984.
- MACHADO, R. J., ROCHA, A. F. *Evolutive fuzzy neural networks*. Trabalho apresentado na International Conference on Fuzzy Systems. San Diego, 1992.
- MORETTIN, P. A., MESQUITA, A.R., ROCHA, J. G. C. Rainfall at Fortaleza in Brazil revisited. In: ANDERSON, O. D., ROBINSON, E. A., ORD, K. (eds.). *Time series analysis: theory and practice 6*. Amsterdam: North-Holland, 1985.
- PORTUGAL, M. S. Neural networks versus time series: a forecasting exercise. *Revista Brasileira de Economia*, v.49, n.4, p.611-629, 1995.
- ROBBINS, H., MONRO, S. A stochastic approximation method. *Annals of Mathematical Statistics*, v.25, p.737-744, 1951.
- RUMELHART, D. E., WEIGEND, S. A. *Predicting the future: a connectionist approach*. Stanford PDP-90-01, 1990 (PARC-SSL-90-20).
- SWANSON, N., WHITE, H. *Can neural networks forecast in the bib leagues? Comparing network forecasts to the pros*. Trabalho apresentado no XIV International Symposium on Forecasting. Estocolmo, 12 a 16 de junho de 1994.
- VARFIS, A., VERSINO, C. *Univariate economic time series forecasting*. Cambridge: Cambridge University Press, 1990.
- WASSERMAN, P. D. *Neural computing: theory and practice*. New York: Van Nostrand Reinhold, 1989.
- WEST, M., HARRISON, J. *Bayesian forecasting and dynamic models*. New York: Springer-Verlag, 1989.
- WHITE, H. *Artificial neural networks: approximation and learning theory*. Oxford: Blackwell Publishers, 1992.

ZANDONADE, E. *Aplicação da metodologia de redes neurais em previsão de séries temporais*. Rio de Janeiro: PUC/Departamento de Engenharia Elétrica, 1993 (Dissertação de Mestrado).

(*Originais recebidos em fevereiro de 1996. Revistos em março de 1996.*)