

REDES NEURAIIS RECORRENTES PARA PREDIÇÃO RECURSIVA DE SÉRIES TEMPORAIS CAÓTICAS: UM ESTUDO COMPARATIVO

ANANDA L. FREIRE*, JOSÉ MARIA P. DE MENEZES JR.*, GUILHERME A. BARRETO*

*Universidade Federal do Ceará, Departamento de Engenharia de Teleinformática
Av. Mister Hull, S/N, CP 6005, CEP 60455-760, Fortaleza, Ceará.

Emails: anandalf@gmail.com, josemenezesjr@gmail.com, guilherme@deti.ufc.br

Abstract— Este artigo tem como objetivo avaliar diferentes arquiteturas de redes neurais recorrentes para predição de séries temporais caóticas. Para este fim, são testados diferentes paradigmas de projeto de redes multicamadas, a saber: rede MLP, rede NARX, rede de Elman e Rede de Ecos de Estados (*Echo-State Networks*). Os desempenhos destas redes são comparados na tarefa de predição multipassos-adiante, também conhecida como predição recursiva, usando a conhecida série do laser caótico.

Keywords— Redes recorrentes, Teorema de Takens, rede NARX, rede de ecos de estado, predição recursiva.

Resumo— This paper reports results on a performance comparison of different recurrent network architectures for chaotic time series prediction. For this purpose, we tested different design paradigms of multilayered neural networks, namely: MLP network, NARX network, Elman network and Echo-state network. Their performances are compared on a multistep-ahead prediction task, also known as recursive prediction, using the well-known chaotic laser time series data.

Palavras-chave— Recurrent networks, Takens' theorem, NARX network, Echo-state network, recursive prediction.

1 Introdução

Redes neurais artificiais (RNAs) têm sido utilizadas com sucesso em tarefas de predição e modelagem de série temporais; ver, por exemplo, Palit & Popovic (2005), para uma pesquisa recente de técnicas e aplicações em diversas áreas. Em particular, quando séries temporais são geradas por um sistema dinâmico não-linear e possuem dependências temporais de longo período (também conhecido como processos de memória longa), modelos de RNA geralmente tem melhor desempenho que as técnicas lineares tradicionais, tais como os modelos de Box-Jenkins. Neste caso, as habilidades de generalização e aproximação universal de funções de RNA justificam seu melhor desempenho preditivo.

Na tarefa de predição de séries temporais um-passo-adiante (UPA), modelos de RNA são usados para estimar somente o próximo valor de uma série temporal, sem realimentar o valor de saída predito para a entrada da rede. Para predição de horizonte mais amplo, faz-se necessário um procedimento conhecido como predição multipassos-adiante (MPA), em que a saída do modelo deve ser realimentada para a entrada de forma recursiva até atingir o instante futuro desejado. A predição MPA é mais difícil de lidar do que a predição UPA e estas são tarefas complexas em que os modelos de RNA desempenham um importante papel, em particular, aqueles relacionados à arquiteturas neurais recorrentes (Principe et al. 2000)

RNAs recorrentes têm laços de realimentação local e/ou global em sua estrutura. Redes *feedfor-*

ward (i.e. sem laços de realimentação) podem ser adaptadas para processar séries temporais através de uma entrada com linha de atrasos com derivações (*tapped delay line*) e treinada pelo algoritmo *backpropagation*, elas podem também ser facilmente convertidas em arquiteturas recorrentes simples por realimentações de ativações dos neurônios das camadas escondidas ou de saída, dando origem às redes de Elman e de Jordan, respectivamente.

Lin et al. (1998) demonstraram que aprender dependências de longo prazo com técnicas do tipo gradiente descendente é mais eficiente numa classe de RNAs conhecida como modelos NARX (*Nonlinear AutoRegressive model with exogenous inputs*) do que em modelos recorrentes de Elman e Jordan. Um procedimento de projeto da rede NARX para aplicação em predição de série temporais univariadas foi proposto em Menezes-Júnior & Barreto (2006a) e Menezes-Júnior & Barreto (2006b). No procedimento proposto não há variável exógena “real”, mas sim uma espécie de variável exógena “virtual”, obtida a partir da série univariada original. De acordo com esta abordagem, existem dois regressores de entrada para a rede NARX. Um deles utiliza a realimentação mais recente da saída rede e o outro é gerado de acordo com o teorema de Takens (1981). O primeiro regressor desempenha o papel de um modelo autorregressivo não-linear, enquanto o segundo desempenha o papel de variável exógena virtual.

Recentemente, um novo paradigma de projeto de redes recorrentes foi proposto sob a alcunha de “Computação de Reservatório” (*Reservoir Compu-*

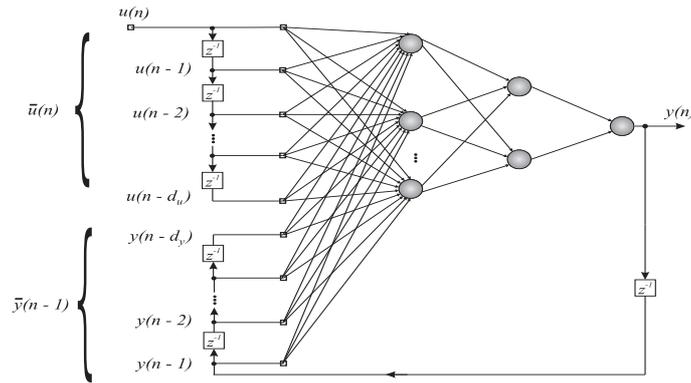


Figura 1: Rede NARX com d_u entradas e d_y atrasos da saída.

ting) (Schrauwen et al. 2007). O interesse neste paradigma por parte da comunidade científica vem aumentando devido principalmente à forma singular de projeto de tais arquiteturas, que não requer o treinamento dos neurônios não-lineares da camada oculta, chamada de reservatório neste contexto. Apenas os neurônios da camada de saída têm seus pesos ajustados, o que pode ser realizado por métodos não-recursivos (e.g. mínimos quadrados). Uma das arquiteturas recorrentes que estão sob o paradigma da computação de reservatório é a rede de Ecos de Estados (*Echo-state network*) (Jaeger 2001).

Do exposto, este artigo tem o propósito de avaliar principalmente os desempenhos das redes NARX e Ecos de Estados comparando-os com os desempenhos de outras arquiteturas neurais, tais como MLP e Elman, na tarefa de predição de predição multipassos-adiante, também conhecida como predição recursiva, usando a conhecida série do laser caótico. O objetivo deste estudo é ressaltar as diferenças no projeto de tais arquiteturas neurais recorrentes, a fim de oferecer subsídios ao usuário no momento da escolha da arquitetura mais adequada à tarefa de interesse.

O restante do artigo esta organizando da seguinte forma. A Seção 2 traz uma apresentação da rede NARX aplicada à predição de séries temporais univariadas. A rede de Ecos de Estados é brevemente descrita na Seção 3. Por fim, na Seção 4 são discutidos os resultados da predição recursiva obtidos pelos vários paradigmas de redes neurais recorrentes, utilizando a série do laser caótico como *benchmarking*.

2 A Rede NARX

Uma importante classe de sistemas não-lineares de tempo discreto é matematicamente representada pelo modelo NARX (*Nonlinear AutoRegressive*

model with eXogenous inputs)

$$y(n) = f[y(n-1), \dots, y(n-d_y); u(n), u(n-1), \dots, u(n-d_u+1)], \quad (1)$$

em que $u(n) \in \mathbb{R}$ e $y(n) \in \mathbb{R}$ representam, respectivamente, a entrada e a saída do modelo no instante n , enquanto $d_u > 0$ e $d_y > 0$, $d_u \leq d_y$, são as ordens dos regressores de entrada e memória de saída.

Assume-se que $f(\cdot)$ é um mapeamento não-linear desconhecida. Quando esta função é aproximada por uma rede MLP, a topologia resultante é chamada de *rede recorrente NARX* (Narendra & Parthasarathy 1990) (Figura 1). Quando a ordem da memória da saída da rede NARX é feita $d_y = 0$, reduz-se a rede NARX a uma rede MLP padrão.

Dada uma série temporal $\{x(n)\}_{n=1}^N$, define-se o regressor do sinal de entrada da rede NARX, denotado por $\mathbf{U}(n)$. Assim, tem-se que

$$\begin{aligned} \mathbf{U}(n) &= [u(n), u(n-1), \dots, u(n-d_u+1)], \quad (2) \\ &= [x(n), x(n-\tau), \dots, x(n-(d_E-1)\tau)], \end{aligned}$$

em que se percebe que o regressor $\mathbf{U}(n) \in \mathbb{R}^{d_u}$ é composto de d_E valores observados da série temporal, amostrados a cada τ unidades de tempo.

Para levar em conta a informação dinâmica fornecida pelo laço de realimentação, a rede NARX pode ser treinada e usada nos modos paralelo e série-paralelo. No modo de treinamento paralelo (ou recorrente), o regressor do sinal de saída, representado por $\mathbf{Y}(n)$, é definido como

$$\mathbf{Y}(n-1) = [y(n-1), y(n-2), \dots, y(n-d_y)], \quad (3)$$

em que se percebe que o regressor de saída $\mathbf{Y}(n) \in \mathbb{R}^{d_y}$ compreende d_y saídas prévias da rede neural. Vale lembrar que, para uma rede treinada adequadamente, a saída da rede no instante n é uma estimativa do valor futuro da série, ou seja, $y(n) = \hat{x}(n+1)$. Assim, a Equação (3) pode ser escrita também da seguinte forma:

$$\mathbf{Y}(n-1) = [\hat{x}(n), \hat{x}(n-1), \dots, \hat{x}(n-d_y+1)]. \quad (4)$$

Já, no modo série-paralelo (ou não-recorrente), o regressor de saída é definido como segue

$$\mathbf{Y}(n-1) = [x(n), x(n-1), \dots, x(n-d_y+1)], \quad (5)$$

sendo, portanto, construído com amostras reais da série temporal de interesse. Qualquer que seja o modo de uso da rede NARX, ambas as arquiteturas implementam o seguinte mapeamento

$$y(n) = \hat{x}(n+1) = \hat{f}[\mathbf{Y}(n-1), \mathbf{U}(n)], \quad (6)$$

em que a função não-linear $\hat{f}(\cdot)$ pode ser realizada prontamente pela rede MLP padrão e treinada pelo algoritmo *backpropagation* convencional.

Um importante destaque a respeito da aproximação proposta é relacionada com a estabilidade da aprendizagem da rede NARX. Em Narendra & Parthasarathy (1990) é mostrado que o modo de treinamento paralelo é muito mais instável do que o modo série-paralelo. Isto se deve basicamente à propagação do erro de predição gerada pelas estimativas realimentadas para o regressor de saída. Desta forma, neste artigo utiliza-se a rede NARX somente no modo de treinamento série-paralelo.

3 A Rede de Ecos de Estados

A rede de Ecos de Estados, doravante referenciada apenas pela sigla em inglês ESN, é uma rede neural recorrente formada, normalmente, por unidades sigmoidais que são conectadas aleatoriamente, de maneira que as únicas partes realmente treinadas são as conexões de saída da rede (Jaeger 2001). Desta forma, a rede, em si, é usada apenas como um reservatório e as conexões de saída são treinadas usando regressão linear.

A Figura 2 apresenta a ideia básica da arquitetura ESN, onde as linhas pontilhadas representam conexões que não são necessariamente utilizadas, e a Figura 3 mostra o diagrama da rede, o qual se diferencia da arquitetura usada, por ter apenas um neurônio na saída, ter um único parâmetro na entrada que é a amostra atual da série, e não possuir a matriz \mathbf{W}_{back} .

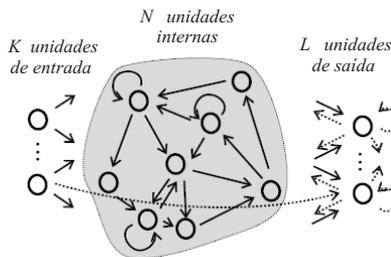


Figura 2: Ilustração da arquitetura ESN.

Na perspectiva dessa rede vê-se que, sob certas condições, o estado de ativação $\mathbf{x}(n)$ de uma rede

neural recorrente é uma função do histórico de entrada $\mathbf{u}(n), \mathbf{u}(n-1) \dots$ apresentada à rede, isto é, existe uma função \mathbf{E} , a qual, $\mathbf{x}(n) = \mathbf{E}(\mathbf{u}(n), \mathbf{u}(n-1))$. Metaforicamente, o estado $\mathbf{x}(n)$ pode ser entendido como um eco da história da entrada (Jaeger 2001).

A arquitetura ESN é uma rede neural de tempo discreto com K unidade de entrada, N unidade internas e L unidades de saída. Ativações das unidades de entrada no instante n são representadas por $\mathbf{u}(n) = [u_1(n) \dots u_K(n)]^T$, e os pesos dessa camada são $\mathbf{W}^{in} = [w_{ij}^{in}]_{N \times K}$. As ativações das unidades internas, denotadas por $\mathbf{x}(n) = [x_1(n) \dots x_N(n)]^T$, são calculadas por meio da seguinte equação

$$\mathbf{x}(n+1) = \mathbf{f} \left(\mathbf{W}^{in} \mathbf{u}(n+1) + \mathbf{W} \mathbf{x}(n) + \mathbf{W}^{back} \mathbf{y}(n) \right) \quad (7)$$

em que $\mathbf{f}(\cdot) = [f_1(\cdot) \dots f_N(\cdot)]^T$ são as funções de ativação das unidades internas, normalmente tangentes hiperbólicas, e os pesos do reservatório são representados pela matriz $\mathbf{W} = [w_{ij}]_{N \times N}$.

As ativações das unidades de saída da rede, representadas por $\mathbf{y}(n) = [y_1(n) \dots y_L(n)]^T$, são dadas por

$$\mathbf{y}(n+1) = \mathbf{f}^{out} \left(\mathbf{W}^{out} \mathbf{z}(n+1) \right), \quad (8)$$

em que $\mathbf{f}^{out}(\cdot) = [f_1^{out}(\cdot) \dots f_N^{out}(\cdot)]^T$ são as funções de ativação das unidades de saída e $\mathbf{z}(n+1) = [\mathbf{u}^T(n+1) \mathbf{x}^T(n+1) \mathbf{y}^T(n)]^T$ é o vetor resultante da concatenação do vetor de entrada atual, do vetor de ativações atuais das unidades internas e do vetor de ativações das unidades de saída no instante anterior. Os pesos da camada correspondente são dispostos na matriz $\mathbf{W}^{out} = [w_{ij}^{out}]_{L \times (K+N+L)}$, e caso haja conexões que se projetam da camada de saída para as unidades internas, a matriz de pesos correspondente é dada por $\mathbf{W}^{back} = [w_{ij}^{back}]_{N \times L}$.

O treinamento é dividido em duas partes: (1) período de amostragem das ativações do reservatório, e (2) cálculo da matriz \mathbf{W}^{out} . Durante a primeira fase, as matrizes \mathbf{W}_{in} , \mathbf{W}_{back} são geradas aleatoriamente e a matriz \mathbf{W} é gerada aleatoriamente de forma esparsa (com 1% a 10% de pesos diferentes de zero), sendo normalizada de forma $\mathbf{W}_{new} = \frac{1}{|\lambda_{max}|} \mathbf{W}_{old}$, onde $|\lambda_{max}|$ é o valor absoluto do maior autovalor da matriz \mathbf{W}_{old} . Em seguida a matriz resultante tem a sua escala mudada pelo raio espectral $\alpha < 1$ em $\mathbf{W} = \alpha \mathbf{W}_{new}$, para que a rede tenha a propriedade de eco de estados¹.

Em seguida, executa-se a rede de $n = 0$ a $n = n_{max}$, iniciando de um estado inicial zero e desprezando a camada de saída. Exclui-se o estado transitório inicial, após o qual os efeitos do estado inicial da rede tenham se esvaído no tempo. Os estados $\mathbf{u}(n)$, $\mathbf{x}(n)$ e $\mathbf{y}(n)$ de $n = n_{min}, \dots, n_{max}$ são ar-

¹Estas heurísticas são sugeridas por Jaeger (2001).

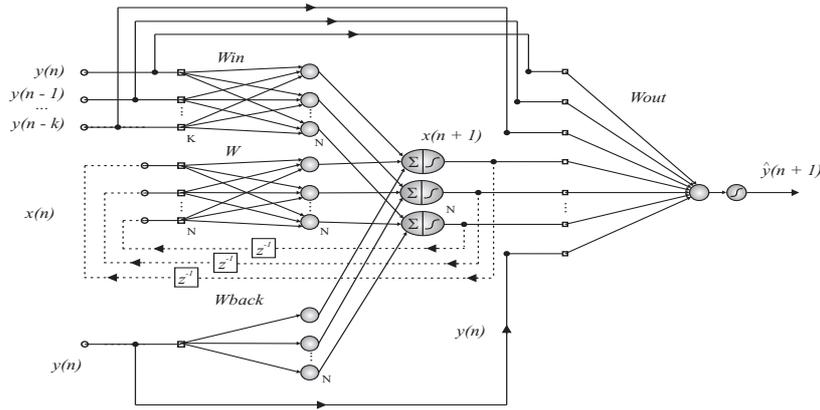


Figura 3: Arquitetura ESN aplicada em previsão de séries temporais univariadas.

mazenados na matriz $\mathbf{M}_{(n_{max}-n_{min}+1) \times (K+N+L)}$ e, ao mesmo tempo, a saída desejada $d(n) = y(n+1)$ é armazenada na matriz $\mathbf{T}_{(n_{max}-n_{min}+1) \times L}$ ².

O último passo de treinamento é a atualização dos pesos da rede. Esta tarefa é simplificada pois é necessário somente otimização dos pesos da camada de saída, representados pela matriz \mathbf{W}^{out} . Desta forma, pode-se utilizar um algoritmo não-recursivo, tal como o método de estimação dos mínimos quadrados, também conhecido como método da pseudoinversa, ou seja

$$(\mathbf{W}^{out})^T = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{T}. \quad (9)$$

4 Simulações e Resultados

A série caótica utilizada para avaliação das redes recorrentes provém de uma sequência de medidas da intensidade de pulsação de um laser de NH_3 infravermelho, obtida de um experimento realizado por Hübner et al. (1989). Esta série temporal foi disponibilizada inicialmente como parte de uma competição de previsão de séries temporais promovida pelo Instituto Santa Fé, ocorrida nos Estados Unidos em 1992. A série temporal utilizada possui 1100 amostras, sendo que as 1000 primeiras são destinadas para o treino e 100 últimas para o teste. Logo, o horizonte de previsão é de 100 unidades de tempo. O procedimento de previsão é recursiva, ou seja, os valores preditos são realimentados para a entrada a fim de compor os regressores de entrada e/ou saída das redes.

As observações da série são normalizadas para o intervalo $[-1, +1]$ e ainda passam pela função tangente hiperbólica. A razão desta última é uma heurística sugerida por Jaeger (2001). São utilizadas 1000 amostras da série para treinamento, das quais as 300 primeiras são eliminadas como amostras pertencentes ao transiente inicial.

²No presente artigo usa-se $L = 1$, logo a matriz \mathbf{T} reduz-se a um vetor.

Para a rede NARX, obteve-se após alguma experimentação os seguintes valores: $d_E = 12$, $\tau = 2$ e $n_y = 28$. O algoritmo *backpropagation* é utilizado para treinar a rede por 3000 épocas, com taxa de aprendizagem fixa e igual a 0,01. Os resultados da previsão são calculados para a média de 10 rodadas de treinamento-teste com reinicializações aleatórias dos pesos da rede.

A arquitetura ESN utilizada tem apenas uma unidade de entrada (i.e. $K = 1$), que consiste na observação atual $y(n)$ da série temporal do laser caótico e um único neurônio na camada de saída ($L = 1$) que fornece a previsão da próxima amostra da série. Após intensa experimentação, o número de unidades do reservatório foi definido em $N = 1000$, com uma densidade de 5% de pesos diferentes de zero e possui um raio espectral $\alpha = 0,99$. As matrizes de pesos possuem valores aleatórios que pertencem ao intervalo $[-0,8 - 0,8]$. Não há realimentação da saída nem para o reservatório e nem para a camada de saída. Tanto a camada oculta, como a de saída, recebem informação apenas da camada de entrada e da ativação dos neurônios do reservatório.

Apesar da camada de saída possuir neurônios com função de ativação tangente hiperbólica, o treinamento ainda foi feito com o mesmo procedimento da Eq. 9, com a exceção do fato da matriz \mathbf{T} armazenar, no lugar da saída desejada, a inversa da função de ativação: $\tanh^{-1}(d(n))$. Como métrica de avaliação do desempenho é utilizado o erro quadrático médio normalizado (NMSE, sigla em inglês).

Na Tabela 1 são listados resultados previamente publicados em Gers (2001) e Bontempi et al. (1999) para a série do laser caótico, tanto para previsão UPA e MPA. Nesta mesma tabela constam também os resultados obtidos no presente trabalho para as redes MLP, Elman, NARX e ESN. Na Figura 4 são mostrados apenas os resultados em ordem de magnitude do NMSE para a previsão recursiva, onde pode-se observar o bom desempenho

das redes NARX e ESN quando comparado com outros modelos de predição. A rede de Elman não apresentou resultados considerados bons em nenhuma das duas tarefas de predição.

Tabela 1: Desempenhos preditivos de várias arquiteturas neurais para a série do laser caótico.

Referência	NMSE UPA	NMSE MPA
LSTM (Gers 2001)	0,3642	0,9683
MLP (Gers 2001)	0,0996	0,8569
FIR-MLP (Wan 1994)	0,023	0,0551
sFIR-MLP (Wan 1994)	-	0,0273
MLP (Koskela et al. 1998)	0,0177	-
RSOM (Koskela et al. 1998)	0,0833	-
EP-MLP (Bakker et al. 2000)	-	0,2159
(Sauer 1994)	-	0,077
MLP (Weigend & Gershefeld 1994)	0,0198	0,016
(Bontempi et al. 1999)	-	0,029
MLP (16-14-10-1)	0,1717	0,7548
Elman (16-14-2-1)	0,0340	0,8479
NARX (12-12-6-1))	0,0339	0,0565
ESN (1000))	0,0657	0,1356

Os detalhes das redes que geraram os resultados da Tabela 1 são apresentados a seguir. Wan (1994) reportou o melhor resultado submetido para a competição promovida pelo instituto Santa Fé. Ele usou uma rede FIR-MLP³ com 25 unidades de entradas e 12 neurônios ocultos. Esta rede modela cada neurônio como um filtro FIR, ou seja, cada neurônio possui memória. A principal consequência desta abordagem é que a rede torna-se difícil de sintonizar e extremamente instável. Para predição recursiva, o autor teve que utilizar observações suavizadas dando origem à rede *smoothed* FIR-MLP. Koskela et al. (1998) compara a rede SOM recorrente (RSOM) com a rede MLP treinada com o algoritmo Levenberg-Marquardt (LM) usando em ambas uma janela de entrada com $d_E = 9$. Bakker et al. (2000) usa uma mistura de valores preditos e valores observados como entrada, a fim de diminuir a propagação de erro devido à realimentação. Além disso, este autor usa PCA (Análise de Componentes Principais) para reduzir a dimensionalidade da entrada de 40 mais recentes valores para 16 componentes principais. Uma rede MLP foi usada para previsão utilizando o algoritmo *backpropagation through time* (BPTT) para treinamento. Na etapa de teste recursivo apenas os valores preditos são realimentados para a entrada.

O melhor resultado para a previsão MPA relatado em Gers (2001) foi o obtido por Weigend & Gershefeld (1994) utilizando uma rede MLP com 25 unidades de entrada, 12 neurônios ocultos e com uma saída adicional para estimar o erro de predição. Bontempi et al. (1999) utiliza métodos de aprendizagem local para a predição recursiva. Finalmente, Gers (2001) utiliza a rede recorrente *Long Short-Term Memory* (LSTM), que possui blocos de memória para guardar informação temporal. Deve-

³FIR = Finite Input Response.

se destacar que o bom desempenho da rede NARX para predição MPA foi obtido usando o algoritmo de treinamento *backpropagation* padrão, enquanto todas as outras redes cujos resultados estão mostrados na Tabela 1 usam métodos de treinamento mais sofisticados (e.g. LM e BPTT).

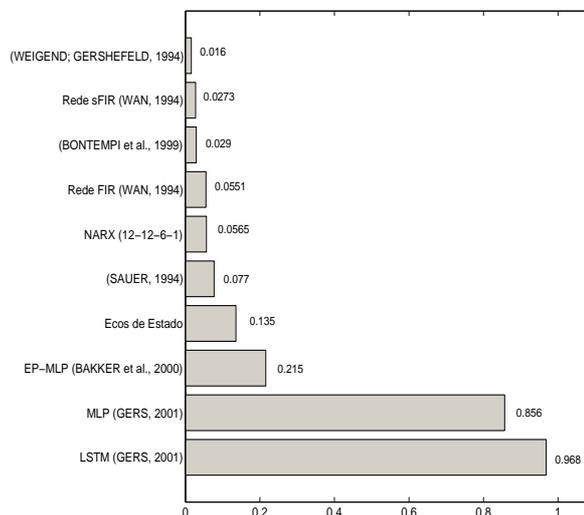


Figura 4: Resultado comparativo entre os modelos para predição MPA.

As Figuras 5 e 6 apresentam os valores preditos recursivamente pela rede NARX e pela arquitetura ESN, respectivamente, para 100 passos adiante. A linha tracejada representa os 100 valores da amplitude estimados recursivamente e a linha sólida são os valores observados da série. É importante salientar que na situação crítica de colapso da intensidade do laser caótico, que ocorre por volta do instante de tempo 60, os dois modelos conseguiram recuperar a dinâmica do sistema mesmo após o colapso. Entre as duas arquiteturas, a rede NARX teve desempenho superior ao do modelo ESN.

Embora os modelos NARX e ESN não tenham apresentado o menor erro de predição para o horizonte de 100 passos, não implica que eles sejam “piores” que os outros modelos comparados. Em testes adicionais, observou-se que a rede NARX e a arquitetura ESN são capazes de reproduzir (modelar) as dinâmicas do laser caótico para horizontes mais longos, melhor do que, por exemplo, o modelo de Wan (1994).

5 Conclusões

Neste trabalho são avaliados os desempenhos dos modelos neurais recorrentes NARX e ESN na predição multipassos-adiante (recursiva) da série do laser caótico. Os resultados obtidos são compara-

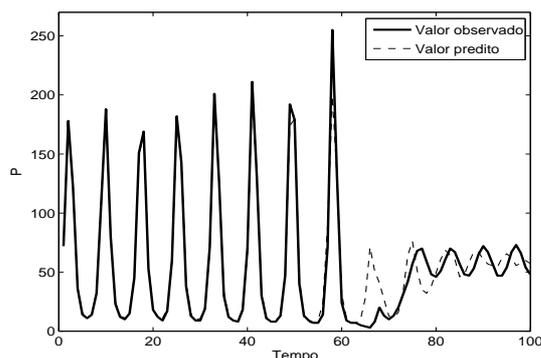


Figura 5: Predição MPA para a série do laser caótico utilizando a rede NARX.

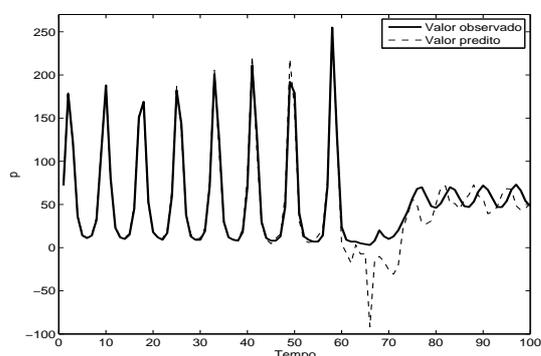


Figura 6: Predição MPA para a série do laser caótico utilizando a arquitetura ESN

dos em um contexto mais amplo com àqueles obtidos por diversas arquiteturas para a mesma tarefa. Uma análise dos resultados obtidos mostrou que os modelos NARX e ESN apresentaram bom desempenho quando comparado com modelos neurais treinados com algoritmos de aprendizagem mais sofisticados, tais como Levenberg-Marquardt (LM) ou *Backpropagation through Time* (BPTT). Em trabalhos futuros pretende-se melhorar ainda mais o desempenho da rede NARX através do uso de outros algoritmos de treinamento, tais como filtro de Kalman estendido, *Real-Time Recurrent Learning* (RTRL) e BPTT (De Jesús & Hagan 2007).

Agradecimentos

Os autores agradecem ao CNPq e FUNCAP pelo suporte financeiro dado a este trabalho.

Referências

Bakker, R., Schouten, J. C., Giles, C. L., Takens, F. C. & Bleek, C. M. V. D. (2000). Learning chaotic attractors by neural networks, *Neural Comput.* **12**(10): 2355–2383.

Bontempi, G., Birattari, M. & Bersini, H. (1999). Local learning for iterated time-series prediction, *ICML '99: Proceedings of the Sixteenth International Conference on Machine Learning*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 32–38.

De Jesús, O. & Hagan, M. (2007). Backpropagation algorithms for a broad class of dynamic networks, *IEEE Transactions on Neural Networks* **18**(1): 14–27.

Gers, F. (2001). *Long short-term memory in recurrent neural networks*, Master's thesis, Universitat Hannover, Hannover.

Hübner, U., Abraham, N. B. & Weiss, C. O. (1989). Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared NH₃ laser, *Physical Review A* **40**: 6354–6365.

Jaeger, H. (2001). The "echo state" approach to analyzing and training recurrent neural networks, *Technical Report GMD Report 148*, German National Research Center for Information Technology.

Koskela, T., Varsta, M., Heikkonen, J. & Kaski, K. (1998). Time series prediction using recurrent som with local linear models, *International Journal of Knowledge-based Intelligent Engineering Systems* **2**: 60–68.

Lin, T., Horne, B. G. & Giles, C. L. (1998). How embedded memory in recurrent neural network architectures helps learning long-term temporal dependencies, *Neural Networks* **11**(5): 861–868.

Menezes-Júnior, J. M. & Barreto, G. A. (2006a). A new look at nonlinear time series prediction with NARX recurrent neural network, *Proceedings of the 9th Brazilian Symposium on Neural Networks (SBRN'2006)*, IEEE Computer Society, pp. 28–33.

Menezes-Júnior, J. M. & Barreto, G. A. (2006b). On recurrent neural networks for auto-similar traffic prediction: a performance evaluation, *Proceedings of the Proceedings of the 2006 IEEE/SBrT International Telecommunications Symposium (ITS'2006)*, pp. 495–500.

Narendra, K. S. & Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks, *IEEE Transactions on Neural Networks* **1**(1): 4–27.

Palit, A. K. & Popovic, D. (2005). *Computational Intelligence in Time Series Forecasting*, 1st edn, Springer Verlag.

Principe, J. C., Euliano, N. R. & Lefebvre, W. C. (2000). *Neural Adaptive Systems: Fundamentals Through Simulations*, John Wiley and Sons.

Sauer, T. (1994). Time series prediction by using delay coordinate embedding, in A. S. Weigend & N. A. Gershenfeld (eds), *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison Wesley, Harlow, UK, pp. 175–193.

Schrauwen, B., Verstraeten, D. & Campenhout, J. (2007). An overview of reservoir computing: theory, applications and implementations, *Proceedings of the 15th European Symposium on Artificial Neural Networks*, pp. 471–482.

Takens, F. (1981). Detecting strange attractors in turbulence, in D. A. Rand & L.-S. Young (eds), *Dynamical Systems and Turbulence*, Vol. 898 of *Lecture Notes in Mathematics*, Springer, pp. 366–381.

Wan, E. A. (1994). *Finite impulse response neural networks with applications in time series prediction*, PhD thesis, Stanford, CA, USA.

Weigend, A. & Gershefeld, N. (1994). *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison-Wesley, Reading.