ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

Simple Reservoirs with Chain Topology Based on a Single Time-Delay Nonlinear Node

J.M. Gutiérrez[1], D. San-Martín[2], S. Ortín[1] and L. Pesquera[1] *

1- Instituto de Física de Cantabria (IFCA). CSIC-Univ. Cantabria
Avda. de los Castros. Santander - Spain

2- Predictia Intelligent Data Solutions, S.L.
CDTUC Fase A, Planta 2 Módulo 203. Santander - Spain

Abstract. A physical scheme based on a single nonlinear dynamical system with delayed feedback has been recently proposed for Reservoir Computing (RC) [1]. In this paper we present a computational implementation of this idea using a simple chain topology with properties derived from its physical counterpart (e.g. the reservoir is defined by two tunable parameters related to feedback- and input-strength terms). An application to time series prediction is described and a comparison with other standard reservoir computing methods is given.

1 Introduction

Standard Reservoir Computing (RC) models are formed by a large reservoir of sparsely and randomly connected neurons (with fixed weights), mapping a dynamical input signal onto a high-dimensional state space in a recurrent and transient-like fashion [2]. Input signals are injected into the reservoir neurons using fixed random weight connections and a simple parametric readout function is trained to combine the resulting reservoir states in order to solve a given task (i.e. the parameters are optimized to reproduce the outputs associated with the inputs). Typically simple linear regression, including ridge, pruning and/or regularization terms [3], are used in the readout process, leading to efficient learning machines.

Due to the "black box" character of RC, the construction of a model for a particular problem requires a series of trials and errors in order to specify optimum connectivity and weight structure of the reservoir for the task, as well as appropriate input connections. However, it has been recently shown that very simple deterministically constructed reservoir with simple cycle topology (e.g. the so-called simple cycle reservoir, SCR) give a performance comparable to standard RC on a number of time series benchmarks [4, 5]. Moreover, due to their simplicity, the properties of these reservoirs can be better analyzed, thus providing more usable machines to solve real world problems.

Along these ideas, a new RC physical paradigm based on a single time-delay nonlinear system has been recently proposed in the electro-optical community [1]. Dynamical time-delay systems have the inherent property that their phase space is infinite-dimensional. Therefore a high-dimensionality can be achieved

*Acknowledgement: The European project PHOCUS (EU FET-Open grant: 240763).

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

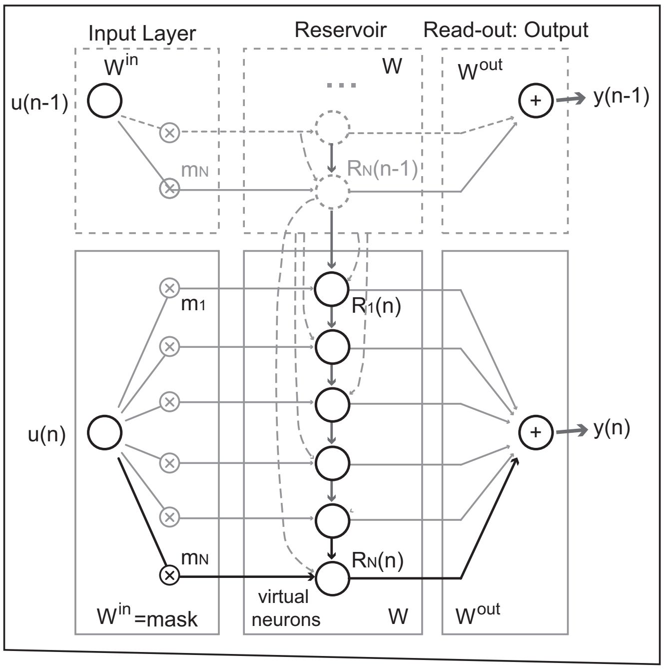

Fig. 1: Schematic overview of the Time-Delayed Reservoir scheme (TDR) with input u, the reservoir $\mathbf{R}$ and the final output y, resulting from the read-out linear function of the reservoir state.

with a single time-delayed system by subdividing the delay time into a number of virtual neurons at equal time intervals. In this way the required complex network of many neurons is reduced to a single system (or node) with delay. In this paper we present a computational implementation of this idea, leading to a simple chain topology similar to the above SCR models, but with neuron functioning and properties derived from its physical counterpart. These models are hereafter referred to as Time-Delayed Reservoirs (TDR) and a schematic illustration of their topology is given in Fig. 1.

2 Time-Delayed Reservoirs (TDR)

The new approach introduced in this paper is based on the following equation:

$$\dot{x}(t) = -x(t) + \rho f[x(t-\tau) + \gamma\, m(t)\, I(t)], \tag{1}$$

characterizing a time-delayed system $x(t)$ with sigmoidal nonlinearity, $f[u] = 1/(1 + exp(-c\,u)) - 0.5$, driven by the input signal $I(t)$. A high-dimensional reservoir of neurons is achieved by subdividing the delay time τ into N virtual nodes at equal time intervals $\Delta = \tau/N$ within τ, all of them receiving the same input signal during the interval $[t, t+\tau]$ but with a fixed mask function $m(t)$,

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

given by a stepwise function with random changes at $t + i\Delta$, $i = 1, \dots, N-1$, taken from a uniform (-1,1) distribution; see [1] for more details.

In this paper we present a computational implementation of the above physical scheme, where each τ time interval is discretized into N sub-intervals and assigned N neurons R_i, $i = 1, \dots, N$, with connections derived from (1) according to the corresponding numerical integration:

$$R_i(n) = (1-h)R_{i-1}(n) + h\,\rho\,f[R_i(n-1) + \gamma\,m_i\,u(n)];\ for\ i \geq 1, \qquad (2)$$

where $R_0(n) \equiv R_N(n-1)$, m_i are the mask values, the parameters γ and ρ correspond to feedback- and input-strength terms, and h corresponds to the integration step (in this paper we use $h = 0.1$). The N different values of m_i are randomly generated from a uniform $(-1, 1)$ distribution and the input data $u(n)$ is standardized.

The above scheme completely defines the topology of the reservoir (a chain), where each neuron is connected to the preceding one in the chain and it is time-delay connected to the same neuron in the previous time (see Fig. 1 for a schematic view). In this case, the activation function f does only affect the latter of these connections, whereas the former one is linear. Finally, note that although the topology of the resulting TDR models seems to be similar to the SCR models [4], Eq. 2 explicitly involves two different reservoir time steps and implies an asynchronous updating of the neurons. Note that this particular topology is inspired in the physical counterpart (1), which required a single time-delayed system to dynamically perform the calculations.

In order to analyze the computing capabilities of the TDR as a function of the parameters, we computed the mean memory capacity using random input data as described in [6]. As it can be shown in Fig. 2, high memory values (up to 16) can be obtained with these type of reservoir in a large area of the parameters, with higher sensitivity to the γ parameter (i.e. the feedback term). Note that this provides a lower limit for the memory achievable by the TDR, since longer memory capacities could be attained for a particular problem, depending on the autocorrelation of the data [7].

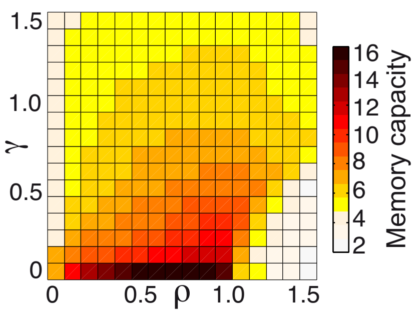

Fig. 2: Memory capacity of the TDR as a function of the γ and ρ parameters.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

3 Time Series Prediction

Time series analysis is a challenging problem with many important real-world applications from engineering to biomedicine and finance [8]. In this paper we consider a particular problem, the Mackey-Glass (MG) time series, which has been previously analyzed in a number of papers comparing the performance of different machine learning methods, including standard reservoir computing [2]. To this aim, we generated a 10000 points time series of the MG system (with standard parameters: $\tau = 17$ and $n = 10$) with equally-spaced sampling every $t = 3$ time units. The first half of the series (5000 points) was used for training and the last one for testing. Fig. 3 shows the results for two particular reservoir sizes: (A) 500 neurons and (B) 1500 neurons. The left column shows the train error (in terms of the normalized $RMSE$, denoted by $NRMSE$, defined as $RMSE/\sigma$, where σ is the standard deviation of training time series) whereas the right one shows the test error, for the same space of ρ and γ parameters considered in Fig. 2. Note that no overfitting occurs in this case and also that areas with smaller errors are associated with those regions with appropriate memory values (see Fig. 2). The optimum results obtained for this time series correspond to the parameter values $\rho = 0.75$ and $\gamma = 0.2$, achieving a $NRMSE$ test error of the order of 10^{-7}. Note that this error corresponds to one-step-ahead prediction, with a sampling step of $t = 3$ and, hence, with a total forecast horizon of $t = 3$.

In order to test the sensitivity of the results to the size of the reservoir and to the particular mask used, we considered a particular TDR configuration ($\rho = 0.6$ and $\gamma = 0.2$) and tested the results for different number of neurons (ranging from 100 to 3000) considering ten different random masks. The bottom panel of Fig. 3 shows the resulting test errors using a box-and-whiskers plot. As it can be shown, the results are not sensitive to the particular mask considered, and optimum results are achieved with around 1500 neurons, as shown in (B).

In order to compare the results with other studies, we have also computed the errors for k steps ahead prediction (with forecast horizons $t = 3\,k$), where the estimated output is re-injected into the reservoir to obtain the next step prediction. Figure 4 shows the results obtained for increasing forecast horizons, ranging from 3 to 2000 (with horizon steps of $t = 30$). In this case, we considered a configuration given by $\rho = 0.6$ and $\gamma = 0.2$, close to the optimum, with a NRMSE of 9.8×10^{-7}. This figure illustrates the good performance of this simple reservoir computing machine; for instance, these results for this particular configuration are close to those reported in [2], with an error of $log_{10}(NRMSE) = -5.09$ for a forecast horizon $t = 84$, claiming an improvement in more than three orders of magnitude with regards to previous results (note that the fourth point shown in Figure 4 corresponds to a forecast horizon of $t = 93$ with $log_{10}(NRMSE) = -4.41$).

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

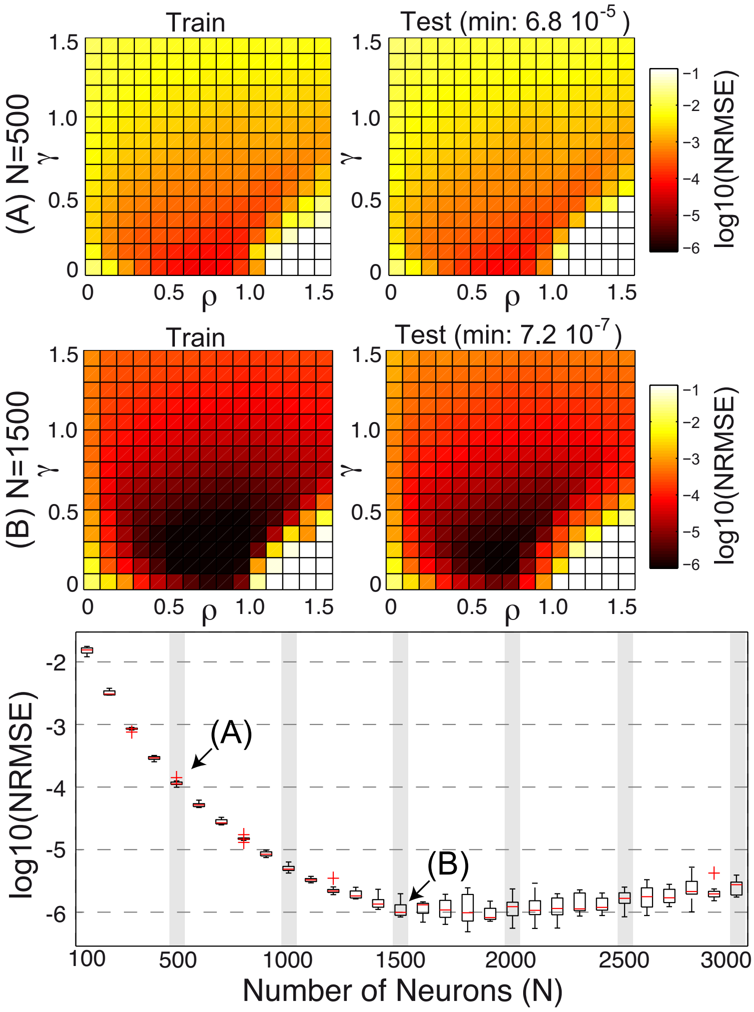

Fig. 3: Results of the performance of the TDR (in terms of normalized RMSE) for two particular reservoir sizes: (A) 500 neurons and (B) 1500 neurons. The bottom figure shows a box-and-whiskers plot of the errors attained for different reservoir sizes (from 100 to 3000 neurons) for the same configuration, $\rho = 0.6$ and $\gamma = 0.2$, and with ten different realizations of random masks.

4 Conclusions

A new physically-inspired reservoir computing scheme has been presented, based on a single nonlinear dynamical system with delayed feedback. This leads to a simple chain topology for the reservoir, which is characterized by two interpretable parameters derived from its physical counterpart: the feedback- and

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

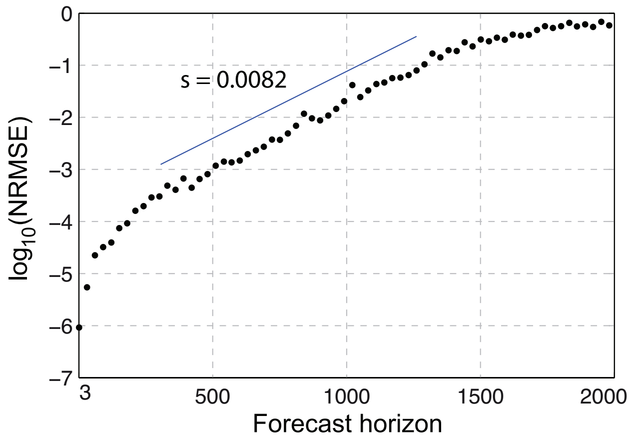

Fig. 4: Test error of the TDR reservoir with parameters $\gamma = 0.2$ and $\rho = 0.6$ for different forecast horizons, ranging from $t = 3$ (one-step-ahead prediction) to $t = 2000$. The slope shows the value of the Lyapunov exponent for MG.

input-strength terms, respectively. The properties of the resulting reservoir are analyzed, both in terms of the memory capacity (with maximum attainable values of 16 for random input) and the prediction error for the reconstruction of time series. In particular, an application to a benchmarking problem (Mackey-Glass time series) is described and a comparison with other standard benchmark reservoir computing methods is given, obtaining similar results.

References

[1] L. Appeltant, M.C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C.R. Mirasso, and I. Fischer, Information processing using a single dynamical node as complex system, *Nature Communications*, DOI 10.1038/ncomms1476, 2011.

[2] H. Jaeger and H. Haas, Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication, *Science*, vol. 304, pp. 78-80, 2004.

[3] X. Dutoit and B. Schrauwen and J. Van Campenhout and D. Stroobandt and H. Van Brussel and M. Nuttin, Pruning and regularization in reservoir computing, *Neurocomputation*, 72, 1534–1546, 2009.

[4] A. Rodan and P. Tiňo, Simple Deterministically Constructed Recurrent Neural Networks. In C. Fyfe et al., editors, proceedings of the 2010 *Intelligent Data Engineering and Automated Learning* (IDEAL 2010), Lecture Notes in Computer Science 6283, pages 267-274, Springer-Verlag, 2010.

[5] A. Rodan and P. Tiňo, Minimum Complexity Echo State Network, *IEEE Transactions on Neural Networks*, 22, 131-144, 2011.

[6] H. Jaeger, Short term memory in echo state networks, German National Research Center for Information Technology Bremen, Tech. Rep. 152, 2002.

[7] M. Hermans and B. Schrauwen, Memory in linear recurrent neural networks in continuous time, *Neural Networks*, 23, 341-355, 2010.

[8] H. Kantz and T. Schreiber. *Nonlinear time series analysis*, Cambridge University Press, Cambridge, 1997.