Neural Networks 83 (2016) 109–120

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Computational analysis of memory capacity in echo state networks

Igor Farkaš*, Radomír Bosák, Peter Gergeľ

Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Mlynská dolina, 84248 Bratislava, Slovakia

ARTICLE INFO

Article history:
Received 13 September 2015
Received in revised form 25 July 2016
Accepted 28 July 2016
Available online 16 August 2016

Keywords:
Echo-state network
Memory capacity
Spectral properties
Reservoir orthogonalization

ABSTRACT

Reservoir computing became very popular due to its potential for efficient design of recurrent neural networks, exploiting the computational properties of the reservoir structure. Various approaches, ranging from appropriate reservoir initialization to its optimization by training have been proposed. In this paper, we extend our previous work and focus on short-term memory capacity, introduced by Jaeger in case of echo state networks. Memory capacity has been previously shown to peak at criticality, when the network switches from a stable regime to an unstable dynamic regime. Using computational experiments with nonlinear ESNs, we systematically analyze the memory capacity from the perspective of several parameters and their relationship, namely the input and reservoir weights scaling, reservoir size and its sparsity. We also derive and test two gradient descent based orthogonalization procedures for recurrent weights matrix, which considerably increase the memory capacity, approaching the upper bound, which is equal to the reservoir size, as proved for linear reservoirs. Orthogonalization procedures are discussed in the context of existing methods and their benefit is assessed.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Memory is a crucial component of any neural system for its functioning in the world. On long timescales, the memories are likely to be stored in synaptic weights. On shorter timescales, the short-term memory (STM) is conjectured to be due to transient network activity. Specifically, stimulus perturbations can cause activity in a recurrent network long after the input has been removed, and the research hypothesizes that cortical networks may rely on transient activity to support STM (Buonomano & Maass, 2009; Jaeger, 2004; Maass, Natschläger, & Markram, 2002). Understanding the role of memory requires determining the limits of STM and characterizing the effects of various network parameters on that capacity (e.g. the network size, topology, and input statistics).[1]

In the domain of recurrent neural networks that possess STM capacity, reservoir computing (RC) (Lukoševičius & Jaeger, 2009) has turned out to be an efficient alternative to computationally demanding gradient-based learning algorithms. RC is based on an appropriate initialization of the input and recurrent part (reservoir) of the network, and only the output connections (readout) are trained in supervised way. RC has been used to quantify the STM capacity in several network architectures, such as spiking networks (Legenstein & Maass, 2007; Maass et al., 2002; Wallace, Hamid, & Latham, 2013), continuous-time networks (Büsing, Schrauwen, & Legenstein, 2010; Hermans & Schrauwen, 2010), and discrete-time networks (Jaeger, 2001; White, Lee, & Sompolinsky, 2004), including the delayed reservoirs (Grigoryeva, Henriques, Larger, & Ortega, 2015). These analyses have shown that even under optimal conditions, the STM capacity (i.e., the length of the stimulus the network is able to recover) scales linearly with the number of nodes in the reservoir.

Significant effort has been devoted to the quantification of the degree to which different inputs lead to different network states (Büsing et al., 2010; Jaeger, 2001; Legenstein & Maass, 2007; Maass et al., 2002; Strauss, Wustlich, & Labahn, 2012; Wallace et al., 2013). This condition should allow the system to recover the original input by inverse computation. In discrete echo state networks (ESNs), the uniqueness of trajectories is guaranteed by the echo state property (ESP) (Jaeger, 2001; Manjunath & Jaeger, 2013) which, however, does not ensure robustness and output computations can be sensitive to small perturbations. A slightly more robust property looks at the conditioning of the reservoir matrix describing how the system acts on an input sequence (Strauss et al., 2012). Strauss et al. (2012) show that the STM capacity still scales linearly with the network size.

One of the factors affecting the network performance regarding the STM is the reservoir connectivity. A number of numerical and

* Corresponding author.
E-mail address: farkas@fmph.uniba.sk (I. Farkaš).

[1] Of course, memory is not the only neural function; there are other complex, spatio-temporal computations, that neural circuits have to perform on the inputs (Dambre, Verstraeten, Schrauwen, & Massar, 2012).

http://dx.doi.org/10.1016/j.neunet.2016.07.012
0893-6080/© 2016 Elsevier Ltd. All rights reserved.

theoretical studies have shown that, in case of randomly connected spiking networks, STM is much longer in the neutrally stable regime than in the converging or diverging regimes (Bertschinger & Natschläger, 2004; Büsing et al., 2010; Legenstein & Maass, 2007; Maass, Legenstein, & Bertschinger, 2005). However, all these studies assumed low connectivity. Recent work has revealed that in case of higher reservoir connectivity, the STM capacity is reduced (Wallace et al., 2013). In contrast, linear ESNs with high connectivities (appropriately normalized) (Büsing et al., 2010) can have relatively large STM capacities (reaching the number of nodes in the network) (Strauss et al., 2012).

The optimal conditions leading to maximized STM have been related to dynamical stability properties of the reservoir. In Bertschinger and Natschläger (2004) it was found that networks were able to exhibit long memories if they operated *near the edge of chaos*, at the critical border between a stable (ordered) and an unstable (chaotic) dynamics regime. The idea that temporal memory is longest near the edge of chaos was confirmed in subsequent studies (Büsing et al., 2010; Legenstein & Maass, 2007; Maass et al., 2005). In the context of ESNs, Jaeger (2001) defined and quantified the STM by *short-term memory capacity* (MC) that measures the network ability to reconstruct the past information from the reservoir on the network output by computing correlations.

In our previous work (Barančok & Farkaš, 2014) we investigated MC in the context of criticality, and assessed it for various input data sets, both random and structured, and showed how the statistical properties of data and various network parameters affect ESN performance. We estimated MC at the edge of chaos (by computing the Lyapunov exponent, LE, explained in Section 2.5) using a nonlinear ESN, in order to compare results with Boedecker, Obst, Lizier, Mayer, and Asada (2012). Since it is difficult to initialize ESNs with required LEs, in this work, we instead manipulate spectral (and other) parameters of ESNs whose effects on MC, and their relationships, have not yet been systematically investigated. To make the analysis simpler, we restrict ourselves to using a random scalar input driving the ESN. We also introduce two reservoir orthogonalization procedures that lead to an almost maximal increase of MC under certain conditions.

The remaining parts of the paper are organized as follows. Section 2 provides more technical details relevant for our work. Section 3 contains results of experiments. Section 4 concludes the paper. Appendix provides mathematical details of the two reservoir orthogonalization procedures.

2. Related background

Here we recollect relevant information related to ESNs and this paper, namely the memory capacity, reservoir initialization, its orthogonalization and estimating reservoir's criticality (dynamical stability).

2.1. Echo state network model

Fig. 1 shows the ESN model with a single input $u(t)$, N reservoir neurons and L output neurons that we consider. Reservoir activations $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t))^\top$ and output activations $\mathbf{y}(t) = (y_1(t), \ldots, y_L(t))^\top$ are updated according to ESN dynamics given by the formulas

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{w}^{\text{in}} u(t) + \mathbf{W}\mathbf{x}(t-1)) \tag{1}$$

$$\mathbf{y}(t) = \mathbf{f}^{\text{out}}(\mathbf{W}^{\text{out}}\mathbf{x}(t)) \tag{2}$$

where $\mathbf{f} : \mathbb{R}^N \to \mathbb{R}^N$ and $\mathbf{f}^{\text{out}} : \mathbb{R}^N \to \mathbb{R}^L$ are suitable activations functions, we use $f = \tanh$ (applied element-wise) and linear readout $\mathbf{f}^{\text{out}} = \mathbf{id}$. $\mathbf{w}^{\text{in}}$ is the input weight vector, $\mathbf{W}$ and

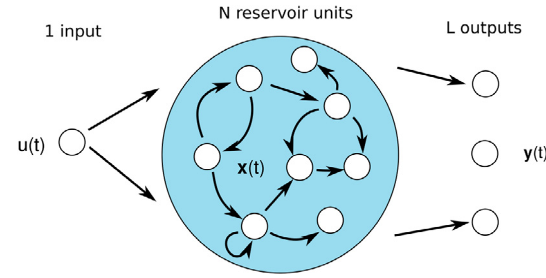

Fig. 1. Illustration of an ESN architecture with a single input.

$\mathbf{W}^{\text{out}}$ are recurrent and output weight matrices, respectively. Readout weights are computed as $\mathbf{W}^{\text{out}} = \mathbf{U}\mathbf{X}^{+}$, where the matrix $\mathbf{U}$ is formed by concatenated desired output vectors (reconstructed past inputs with various delays), and $\mathbf{X}^{+}$ is the pseudoinverse matrix of concatenated state vectors.

2.2. Memory capacity

Jaeger (2001) introduced (short term) memory capacity (MC), as a measure for the ability of the reservoir to store and recall previous inputs fed into the network. Jaeger defined it as

$$\mathrm{MC} = \sum_{k=1}^{k_{\max}} \mathrm{MC}_k = \sum_{k=1}^{k_{\max}} \frac{cov^2(u(t-k), y_k(t))}{var(u(t)) \cdot var(y_k(t))} \tag{3}$$

where cov denotes covariance (of the two time series), var means variance, $k_{\max} = \infty$, $u(t-k)$ is the input presented k-steps before the current input, and $y_k(t) = \mathbf{w}_k^{\text{out}}\mathbf{x}(t) = \tilde{u}(t-k)$ is its reconstruction at the network output (using linear readout), where $\mathbf{w}_k^{\text{out}}$ is the weight vector of kth output unit. The computation of MC is approximated using $k_{\max} = L$ (i.e. given by the number of output neurons). The concept of MC is based on network's ability to retrieve the past information (for various delays k) from the reservoir using the linear combinations of reservoir unit activations (which is quantified by MC_k). The reconstructed past inputs are computed at ESN output.[2]

Jaeger computed MC in ESNs assuming also neural connections between input and output neurons. Following Boedecker et al. (2012), where a number of experiments with MC were performed, we did not assume direct connections, in order to be able to compare the results. The difference in definition does not, however, render both approaches incompatible, since one definition can be seen as a special case of the other. A reservoir with $(N+1)$ units with the last neuron serving as a mere delay of the input can mimic the behavior of a reservoir which allows direct input–output connections. This means that in case of Boedecker's definition, the maximum MC is $N-1$ (or $N - \mathrm{MC}_{k=0} \approx N-1$ to be precise). Jaeger (2001) proved that the memory capacity for recalling an i.i.d. (independent, identically distributed) input by an N-unit ESN with identity activation function is bounded by N.

2.3. Reservoir initialization

Memory capacity obviously depends on the reservoir properties. Lukoševičius (2012) and Lukoševičius and Jaeger (2009) provide a nice and clear overview of practical tips on reservoir initialization in ESNs. The authors suggested that the generated reservoirs be big (to ensure many input signal transformations), sparse (to enforce loose coupling between activation

[2] Defining MC_k as a squared correlation coefficient implies that negative correlations (due to negated reconstructions of the input signal, which is symmetric around zero) are equally considered to contribute to memory capacity.

Download English Version:

https://daneshyari.com/en/article/4946732

Download Persian Version:

https://daneshyari.com/article/4946732

Daneshyari.com