Neurocomputing 000 (2018) 1–12

Contents lists available at ScienceDirect

Neurocomputing

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

Local Lyapunov exponents of deep echo state networks

Claudio Gallicchio*, Alessio Micheli, Luca Silvestri

Department of Computer Science, University of Pisa, Largo Bruno Pontecorvo 3, Pisa, 56127 Italy

A R T I C L E I N F O

Article history:
Received 9 July 2017
Revised 27 October 2017
Accepted 21 November 2017
Available online xxx

Keywords:
Reservoir Computing
Echo state network
Deep learning
Deep recurrent neural networks
Deep echo state network
Stability analysis
Lyapunov exponents

A B S T R A C T

The analysis of deep Recurrent Neural Network (RNN) models represents a research area of increasing interest. In this context, the recent introduction of Deep Echo State Networks (DeepESNs) within the Reservoir Computing paradigm, enabled to study the intrinsic properties of hierarchically organized RNN architectures.In this paper we investigate the DeepESN model under a dynamical system perspective, aiming at characterizing the important aspect of stability of layered recurrent dynamics excited by external input signals.To this purpose, we develop a framework based on the study of the local Lyapunov exponents of stacked recurrent models, enabling the analysis and control of the resulting dynamical regimes. The introduced framework is demonstrated on artificial as well as real-world datasets. The results of our analysis on DeepESNs provide interesting insights on the real effect of layering in RNNs. In particular, they show that when recurrent units are organized in layers, then the resulting network intrinsically develops a richer dynamical behavior that is naturally driven closer to the edge of criticality. As confirmed by experiments on the short-term Memory Capacity task, this characterization makes the layered design effective, with respect to the shallow counterpart with the same number of units, especially in tasks that require much in terms of memory.

1. Introduction

The extension of deep learning methodologies to the class of Recurrent Neural Networks (RNNs) is currently stimulating an increasing interest in the machine learning community [1,2]. In this area, the study of hierarchically structured RNN architectures (see e.g. [3–8]) paved the way to the design of models able to develop feature representations of temporal information at increasing levels of abstraction, enabling a natural approach to tasks on time-series featured by multiple time-scales (especially in the cognitive area). Besides, the elaboration of temporal information in a layered and recurrent fashion is also motivated by strong evidences of biological plausibility emerged from the area of neuroscience [9,10].

However, the analysis of deep RNNs is relatively young, and one of the major topics still deserving research attention is related to understanding and characterizing their dynamical behavior, especially in relation to the inherent role of the hierarchical composition of the recurrent units in layers. A useful methodology in this regard is provided by the Reservoir Computing (RC) [11,12] paradigm and the Echo State Network (ESN) [13,14] approach to RNNs modeling. In particular, allowing to taking apart all the effects due to learning, the recent introduction of the DeepESN model [15,16] enabled the study of the intrinsic role played by the layering factor in deep RNN architectures. Moreover, by inheriting the training characterization typical of standard RC models, DeepESNs also provide an efficient methodology for designing and training deep learning models in the temporal domain.

A first mean to investigate the characteristics of recurrent network dynamics is given in the RC area by the Echo State Property (ESP) [17], which has recently been extended to the case of deep networks in [18]. The analysis provided by the study of the ESP conditions in [17] has started to reveal the natural characterizations of deep RNNs under a dynamical system perspective [18], but it might result of reduced utility in practical cases as it basically neglects the influence of the external input on networks dynamics. By their very nature, recurrent neural models implement dynamical systems whose trajectories in the state space are influenced by initial conditions and by the external input signals, which practically realize a link between the system dynamics and the computational task at hand. In this context, the analysis of stability of deep RNNs dynamics when driven by an external input represents a topic of great importance and still demanded in literature.

In this paper, by pursuing the study of the dynamical behavior of recurrent models typical in the RC area, we provide a theoretical and practical tool that allows us to investigate and control the stability of deep recurrent networks driven by the input. Specifically, we extend the applicability of the study of

* Corresponding author.
E-mail addresses: gallicch@di.unipi.it (C. Gallicchio), micheli@di.unipi.it (A. Micheli), silvestriluca@hotmail.it (L. Silvestri).

https://doi.org/10.1016/j.neucom.2017.11.073
0925-2312/

local Lyapunov exponents [19,20] from the case of shallow ESNs (see e.g. [12,21,22]) to the case of DeepESNs. In particular, the maximum among the local Lyapunov exponents is a useful mean to express the network's sensibility to small perturbations of its state trajectories, and as such it can well quantify the degree of stability (or order) in the dynamical behavior of the system. Given the actual input for the system, the proposed methodology can be used to identify the different dynamical regimes that follow from different cases of networks design conditions, such as the RC scaling factors, the number of recurrent units and the depth of the network. The proposed tool is practically demonstrated on artificial data as well as on signals from real-world datasets.

While the developed tool could be certainly applied to the case of deep RNNs at any stage of training, its application in the RC context enables us to investigate the actual role of layering in RNNs and shed light on its natural effect on the richness and stability of the developed network's dynamics. In this regard, a particularly interesting condition of dynamical behavior is represented by the stable-unstable transition where the maximum local Lyapunov exponent is null, a region of the state space known as the edge of criticality. Previous works in the RC literature already showed that the performance of recurrent models for tasks requiring a long memory span peaks near the criticality of their dynamical behavior [23–26]. Examples are represented by the benchmark tasks in the RC area (e.g. [12,27–30]), tasks in the domain of neural circuit models (e.g. [24,25,31]), as well as real-world tasks, e.g. in the area of speech processing [12] and mobile traffic load estimation [32]. Although the methodology proposed in this paper is not put forward as a performance predictor for trained recurrent models, as an additional element of analysis here we use it to study the relation between the memory and the regimes of DeepESN behaviors through the short-term Memory Capacity task [33].

The rest of this paper is organized as follows. In Section 2 we introduce the basic elements of RC and describe the DeepESN model. In Section 3 we provide the mathematical characterization of the stability analysis of DeepESNs in terms of the maximum local Lyapunov exponent. The outcomes of our experimental analysis are reported and discussed in Section 4. Finally, conclusions are presented in Section 5.

2. Deep echo state networks

Within the framework of randomized neural networks [34], the RC paradigm [11,12] has attested as a state-of-the-art methodology for efficient RNN modeling. The most widely known model in this context is represented by the ESN [13,14,35]. From the architectural perspective, an ESN comprises a recurrent hidden layer of non-linear units, called reservoir, and a feed-forward output layer of typically linear units, called readout. The essence of the ESN operation is that the reservoir part implements a set of randomized filters that serve to dynamically and non-linearly encode the input history into a high dimensional state space, where the task at hand can be approached satisfactorily even by a linear output tool.

From a dynamical system point of view, the reservoir of an ESN computes a discrete-time input-driven non-linear dynamical system, such that at each time step the state evolution is ruled by the reservoir state transition function. By referring to the case of leaky integrator reservoir units [36], at each time step t the reservoir state update equation is given by:

$$\mathbf{x}(t) = (1-a)\mathbf{x}(t-1) + a\tanh(\mathbf{W}_{in}\mathbf{u}(t) + \boldsymbol{\theta} + \hat{\mathbf{W}}\mathbf{x}(t-1)), \quad (1)$$

where $\mathbf{x}(t) \in \mathbb{R}^{N_R}$ and $\mathbf{u}(t) \in \mathbb{R}^{N_U}$ are respectively the reservoir state and the input at time step t, $a \in [0, 1]$ is the leaking rate parameter, $\mathbf{W}_{in} \in \mathbb{R}^{N_R \times N_U}$ is the input weight matrix, $\boldsymbol{\theta} \in \mathbb{R}^{N_R}$ is the weight vector corresponding to the unitary input bias, $\hat{\mathbf{W}} \in \mathbb{R}^{N_R \times N_R}$ is the recurrent reservoir weight matrix and tanh denotes the element-wise application of the hyperbolic tangent non-linearity. Typically, a null state is used as initial condition, i.e. $\mathbf{x}(0) = \mathbf{0}$.

The output at time step t is computed by the readout as a linear combination of the activation of the reservoir units, according to the following equation:

$$\mathbf{y}(t) = \mathbf{W}_{out}\mathbf{x}(t) + \boldsymbol{\theta}_{out}, \quad (2)$$

where $\mathbf{y}(t) \in \mathbb{R}^{N_Y}$ is the output at time step t, $\mathbf{W}_{out} \in \mathbb{R}^{N_Y \times N_R}$ is the output weight matrix and $\boldsymbol{\theta}_{out} \in \mathbb{R}^{N_Y}$ is the vector of weights corresponding to the unitary input bias for the readout.

A major peculiarity of the ESN approach is that only the readout undergoes a training process, such that the weights in $\mathbf{W}_{out}$ and $\boldsymbol{\theta}_{out}$ are adjusted on a training set in order to solve a least squares problem, typically in an off-line fashion and in closed form, using of pseudo-inversion or Tikhonov regularization. The reservoir's parameters are instead left untrained after initialization constrained to the dictates of the Echo State Property (ESP) [14]. The ESP states that the reservoir's dynamics should asymptotically depend only on the driving input signal, while dependencies on initial conditions should vanish with time such that the state of the network tends to represent an "echo" of the input. Essentially, the ESP links the asymptotic behavior of the reservoir dynamics to the input signal on which the reservoir is running. Although a certain research effort has being devoted in the last years to describe and understand more and more in depth the conditions under which the ESP holds (see e.g. [17,37,38]), two basic conditions are widely adopted in literature for this purpose. Specifically, a sufficient condition and a necessary condition are applied to the weight matrix $\hat{\mathbf{W}}$, requiring to respectively control its 2-norm (i.e. its maximum singular value) and its spectral radius (i.e. the maximum among the eigenvalues in modulus) to be below unity. In the following, we will refer to the standard ESN model, as described by Eq. (1) and (2) as *shallow ESN*.

In this paper we are concerned with the extension of the shallow ESN model towards a deep architecture, in which the recurrent component is hierarchically organized into a stack of reservoir layers. The corresponding model is termed DeepESN, as introduced in [15,16]. From a general perspective, it is worth to note that, although several possible ways of constructing deep recurrent architectures have been investigated in literature [3], a stacked composition of recurrent hidden layers is likely to represent the most common choice (see e.g. [4–6,8]).

Focusing on the recurrent part of the architecture of a DeepESN, graphically illustrated in Fig. 1, at each time step the state computation follows a pipeline from the external input towards the higher layer. Specifically, at time step t the first layer is fed by the external input, whereas each layer in the hierarchy at depth higher than 1 is fed by the output of the previous layer at the same time step t.

Keeping the basic notation introduced above for shallow ESNs, here we use N_L to denote the number of reservoir layers in the stacked architecture, assuming for the ease of presentation that every layer has the same dimension (i.e. the same number of recurrent units), which we indicate by N_R. Moreover, for every $i = 1, 2, \dots, N_L$, we use $\mathbf{x}^{(i)}(t) \in \mathbb{R}^{N_R}$ to indicate the state of the reservoir in the ith layer at time step t.

Viewing the DeepESN as a whole system, the global state space of the network can be considered as the product of the N_L state spaces of the layers in the architecture. Accordingly, the global state of the DeepESN at time step t is represented by $\mathbf{x}_g(t) = (\mathbf{x}^{(1)}(t), \mathbf{x}^{(2)}(t), \dots, \mathbf{x}^{(N_L)}(t)) \in \mathbb{R}^{N_L N_R}$. From a dynamical system point of view, the global dynamics of a DeepESN is ruled by its global state transition function F:

$$F : \mathbb{R}^{N_U} \times \underbrace{\mathbb{R}^{N_R} \times \ldots \times \mathbb{R}^{N_R}}_{N_L \text{ times}} \rightarrow \underbrace{\mathbb{R}^{N_R} \times \ldots \times \mathbb{R}^{N_R}}_{N_L \text{ times}} \quad (3)$$