Neurocomputing 138 (2014) 297–309

Contents lists available at ScienceDirect

Neurocomputing

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

Self-organization and lateral interaction in echo state network reservoirs

Levy Boccato*, Romis Attux, Fernando J. Von Zuben

Department of Computer Engineering and Industrial Automation (DCA), FEEC, University of Campinas, 13083-852 Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 7 June 2013
Received in revised form
21 January 2014
Accepted 26 January 2014
Communicated by Manoj Kumar Tiwari
Available online 15 February 2014

Keywords:
Echo state networks
Reservoir computing
Self-organization
Lateral interaction
Channel equalization
Chaotic prediction

ABSTRACT

Echo state networks (ESNs) are recurrent structures that give rise to an interesting trade-off between achievable performance and tractability. This is a consequence of the fact that the key element of these networks – the recurrent intermediate layer known as dynamical reservoir – is not, as a rule, subject to supervised training, which is restricted to the linear output layer, also termed as readout. This trade-off, aside from being of theoretical significance, establishes ESNs as most attractive tools for both online and offline information processing. There are two key aspects to be taken into account in the ESN design: (i) the unsupervised definition of the synaptic weights of the reservoir and (ii) the definition of the structure and of the training strategy associated with the readout. This work is concerned with the first of these aspects: it proposes novel strategies for ESN reservoir design based on the theoretical framework built by Kohonen's classical works on self-organization – which includes the notions of short-range positive feedback and lateral inhibition – and also on the related and more recent notion of neural gas. It is shown, with the aid of a representative set of simulation results, that the proposed methodologies are capable of leading to significant performance improvements in the context of relevant information processing tasks – channel equalization and chaotic time series prediction – particularly when the input data suits well a cluster-based profile.

1. Introduction

Recurrent neural networks (RNNs) can be safely regarded as powerful processing structures in view of a number of features: (i) ability to deal with time context information, (ii) capability of approximating dynamical systems with arbitrary accuracy [11,33] and (iii) presence of feedback connections, which leads to the emergence of a useful dynamical memory of the input signal history [13]. However, the image of a double-edged sword can be used as a portrait of RNNs since the aforementioned attractive characteristics are usually accompanied with well-known drawbacks associated with conventional training strategies, such as slow convergence and instability.

In 2001, a new approach has brought an interesting alternative to circumvent the RNN training difficulties. The proposed model, known as echo state network (ESN) [15], is characterized by the use of an RNN, called the dynamical reservoir, whose parameters – input and recurrent synaptic weights – are randomly created, and of an adaptive readout, which produces the network outputs by means of linear combinations of the reservoir activations. By keeping the recurrent layer parameters fixed, the network training consists in determining the optimum coefficients of the linear combiner at the output, which essentially amounts to the solution of a linear regression problem. Hence, not only do ESNs explore, to a certain extent, the advantages of a recurrent structure, but also introduce a significant simplification in the RNN training process [24].

ESNs, along with the so-called liquid state machines (LSMs), proposed by Maass et al. [25], established a new research area known as reservoir computing (RC) [24,40], which is rapidly attracting interest especially because of the promising results such models have achieved in different applications, such as system identification [16] and nonlinear signal processing [17,5]. Additionally, there are potential analogies between RC principles and structural/dynamical properties of mammalian brains [44].

The twofold architecture of ESNs has promoted the development of distinct research lines. The first line focuses on the ESN output layer and aims at proposing alternative readout structures that are capable of improving the accuracy in the approximation of the target signal. For instance, Boccato et al. [6] replaced the linear combiner with the structure of a Volterra filter, along with a compression stage based on Principal Component Analysis (PCA), with the purpose of exploiting the higher-order statistics of the signals generated by the reservoir, and achieved significant performance improvements in the context of signal processing tasks [7]. Another relevant contribution was brought by

* Corresponding author: Tel.: +55 19 35213857.
E-mail address: lboccato@dca.fee.unicamp.br (L. Boccato).

http://dx.doi.org/10.1016/j.neucom.2014.01.036
0925-2312

Butcher et al. [9], which proposed the use of extreme learning machines (ELMs) at the ESN readout.

On the other hand, the second research stream is devoted to studying the effects of the reservoir characteristics on the network performance and to developing alternative methods for the design and the adaptation of the recurrent layer. In this context, the dilemma is to create a sufficiently rich repertoire of dynamical behaviors without violating the spirit of simplicity inherent to the ESNs. The original idea of creating a random sparse weight matrix meets the requirements of low computational complexity and allows the formation of an internal memory of the recent input history [15]. However, it seems intuitive that an especially tailored reservoir should lead to a better performance than a general random procedure. Therefore, recent works, among which we highlight [38,35,8], have investigated the possibility of incorporating relevant information with respect to the input signal into the reservoir design process.

In this work, we propose a novel unsupervised method for designing ESN reservoirs characterized by (i) the introduction of short-range positive lateral feedback between neighboring units, as well as inhibitory stimuli between distant units, and (ii) the self-organization of the input weights. Inspired by the seminal work of Kohonen [19], the recurrent connections are interpreted as promoting lateral interactions between the reservoir neurons and can be modeled according to the profile of the so-called Mexican hat function so that each activated neuron stimulates its neighboring units and, at the same time, inhibits the activation of more distant units. This idea encourages the formation of groups of neurons, or activity clusters, specialized in responding to different classes of input patterns.

In this context, self-organizing the input weights of the reservoir units emerges as a natural complementary approach that contributes to the stable formation of such activity clusters. Hence, the activation of the reservoir neurons may contain relevant information about the input signal, while, implicitly, a certain degree of diversity is maintained due to the existence of different activity clusters.

With respect to the adaptation of the input weights, two different approaches are considered here: self-organizing maps (SOMs) [20] and the neural gas network (NG) [27,10]. Interestingly, in the latter case, since the NG automatically creates and updates a connectivity matrix, we can use it as the weight matrix of the reservoir by applying a proper scaling factor.

All these possibilities are studied in the context of two relevant information processing tasks – channel equalization and chaotic time series prediction – which have been selected due to their different characters regarding the existence of clustered input patterns: in the former task, the input patterns are distributed in separate clusters, whereas in the latter this behavior does not usually occur. The proposed method is compared with usual strategies for the reservoir design, and we also analyze it from the perspective of the formation of activity clusters within the reservoir.

This paper is organized as follows: Section 2 describes the main aspects of echo state networks, along with a brief review of different reservoir design methods. The main elements introduced by Kohonen [19] that motivated this work, as well as the fundamental ideas of the proposed method, are presented in Section 3. Then, Sections 4 and 5 bring the description of the self-organizing methods and the problems – channel equalization and chaotic time series prediction – considered in this work, respectively. Next, Section 6 discusses the results obtained with the ESNs, outlining the potential advantages of the proposed method. Finally, concluding remarks and future perspectives are presented in Section 7.

2. Echo state networks

The basic ESN architecture, depicted in Fig. 1, consists of a recurrent layer of nonlinear processing elements (NPEs) followed by a feedforward structure, usually a linear combiner, which is responsible for combining the signals generated by the reservoir NPEs to produce the network outputs. The first part of the architecture, commonly named as dynamical reservoir, engenders a repertoire of dynamical behaviors, which are influenced by the current network inputs and by past activations of the NPEs that are fed back into the reservoir.

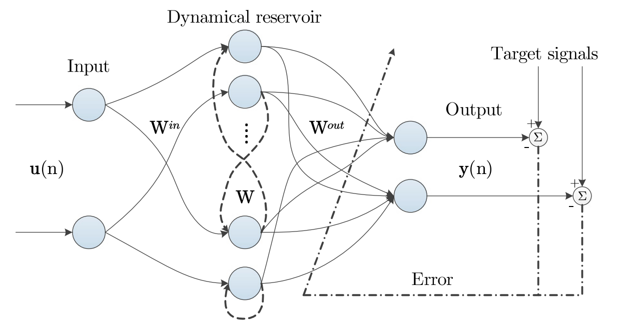

Fig. 1. Basic architecture of an ESN. $\mathbf{W}^{in}$ and $\mathbf{W}$ specify the synaptic weights of the input and recurrent connections, respectively. Only the output weights ($\mathbf{W}^{out}$) are adapted according to the information brought by a target signal.

Consider a generic discrete-time ESN with K input units, N reservoir units and L outputs. The input stimuli, represented by vector $\mathbf{u}(n) = [u_1(n) \ldots u_K(n)]^T$, are linearly combined according to the weights specified in matrix $\mathbf{W}^{in} \in \mathbb{R}^{N\times K}$ and transmitted to the dynamical reservoir, which is composed of fully connected nonlinear neurons whose activations, given by $\mathbf{x}(n) = [x_1(n) \ldots x_N(n)]^T$, represent the network states and are updated as follows [15]:

$$\mathbf{x}(n+1) = \mathbf{f}(\mathbf{W}^{in}\mathbf{u}(n+1) + \mathbf{W}\mathbf{x}(n)), \tag{1}$$

where $\mathbf{W} \in \mathcal{R}^{N\times N}$ contains the weights of the recurrent connections within the reservoir and $\mathbf{f}(\cdot) = (f_1(\cdot), \ldots, f_N(\cdot))$ denotes the activation functions of the internal units.

Then, the network outputs, represented by the vector $\mathbf{y}(n) = [y_1(n) \ldots y_L(n)]^T$, are determined according to the following expression [15]:

$$\mathbf{y}(n+1) = \mathbf{W}^{out}\mathbf{x}(n+1), \tag{2}$$

where $\mathbf{W}^{out} \in \mathcal{R}^{L\times N}$ corresponds to the output weight matrix.

The essential idea underlying the ESN approach is that the parameters of the recurrent layer can be set in advance and independent of the network adaptation, which means that only the coefficients of the readout (elements of matrix $\mathbf{W}^{out}$) are effectively adjusted with the aid of a reference signal. Moreover, due to the linear character of the output layer, the optimum readout parameters can be determined in the least-squares sense by means of linear regression methods [15].

This noticeable simplification in the training process can be brought to fruition due to the so-called echo state property (ESP) [15], which ensures that the activation of each neuron within the reservoir becomes a nonlinear transformation of the recent history of the input signal (hence the term *echo*) as long as the reservoir weight matrix $\mathbf{W}$ meets specific spectral requirements. A sufficient condition for the existence of echo states is expressed in terms of the largest singular value of the internal weight matrix $\mathbf{W}$, which must be smaller than one[1] [15]. Recently, Yuldiz et al. [45] presented a different sufficient condition for the existence of echo states that evokes Schur's definition of matrix stability.

Nevertheless, apart from those boundary conditions, a fundamental issue still needs to be addressed: the design of the

[1] This condition has been demonstrated considering an ESN without output feedback and with tanh(·) as the activation function of the neurons at the reservoir.