

Input Routed Echo State Networks

Luca Argentieri, Claudio Gallicchio and Alessio Micheli *

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

Abstract. We introduce a novel Reservoir Computing (RC) approach for multi-dimensional temporal signals. Our proposal is based on routing the different dimensions of the driving input towards different dynamical sub-modules in a multi-reservoir architecture. At the same time, controllable interconnections among the sub-modules allow modeling the interplay between the different dynamics that might be required by the task. Experiments on synthetic and real-world time-series classification problems clearly show the advantages of the proposed approach in dealing with multi-dimensional signals in comparison to standard RC neural networks.

1 Introduction

Reservoir Computing (RC) [1] is the leading edge approach for efficient learning in temporal domains. Essentially, it is a design methodology for Recurrent Neural Networks (RNNs), and consists in limiting the training algorithms to operate only on the final readout layer of the architecture. As regards the internal connections of the hidden recurrent layer, the reservoir, these are kept fixed after a dynamics-informed initialization, which enable the development of meaningful and rich temporal representation of the driving input signal. Over the years, RC has proved extremely useful in several application contexts, with edge-AI and neuromorphic computing being two outstanding examples.

In this paper, we deal with the problem of designing RC networks in the presence of multi-dimensional input signals. This is a known problem for reservoirs [2], in which the different input dimensions naturally tend to overlap and mix together, resulting in dynamics that might fail to properly represent the specificities and relevance of the different components of the driving signal. We propose a multiple reservoir architecture in which the different dimensions of the input are routed to dedicated sub-reservoirs, which can in this way develop dynamics that follow more closely the evolution of the corresponding one-dimensional signal. The sub-reservoirs are then linked to each other by interconnections whose controllable strength allows the relevant interaction of the input-driven dynamics in a task dependent way. We show the potentialities of the introduced input routed RC on both synthetic and real-world tasks for time-series classification.

*This work has been partially supported by the project TEACHING, under the European Union's Horizon 2020 Research and Innovation program (G.A. ID: 871385), and by the project BrAID, under the Regione Toscana Bando Ricerca Salute 2018 (Decreto n. 15397, 26/09/2018).

2 Input Routed Reservoirs

We first introduce the standard RC approach by making use of the formalism of the Echo State Network (ESN) [3] model. This is an RNN architecture where only the output *readout* layer is adapted on a training set, leaving the hidden recurrent *reservoir* layer unadapted after wise initialization.

Let us consider an N -dimensional input evolving in discrete time, i.e., $\mathbf{x}(t) \in \mathbb{R}^N$, feeding a reservoir layer with H neurons. Starting at the origin, the state of the reservoir, i.e., $\mathbf{h}(t) \in \mathbb{R}^H$, evolves according to the following iterated map:

$$\mathbf{h}(t) = (1 - \alpha) \mathbf{h}(t - 1) + \alpha \tanh(\mathbf{W}_h \mathbf{h}(t - 1) + \mathbf{W}_x \mathbf{x}(t) + \mathbf{b}), \quad (1)$$

where $\alpha \in (0, 1]$ is the leaking rate, $\mathbf{W}_h \in \mathbb{R}^{H \times H}$ is the recurrent weight matrix, $\mathbf{W}_x \in \mathbb{R}^{H \times N}$ is the input weight matrix, $\mathbf{b} \in \mathbb{R}^H$ is a bias vector, and $\tanh(\cdot)$ is the element-wise applied non-linear activation function. Fig. 1(a) graphically shows the ESN architecture. All the parameters of the state transition function in eq. 1 are left untrained, provided that their initialization is in agreement with asymptotic stability of the resulting dynamical system. [4]. In practice, this consists in controlling the spectral radius of the recurrent weight matrix \mathbf{W}_h , typically to values not far beyond unity. Accordingly, the values in \mathbf{W}_h are drawn from a uniform distribution on $[-\rho \sqrt{3/H}, \rho \sqrt{3/H}]$, which determines a spectral radius of ρ [5]. \mathbf{W}_x and \mathbf{b} are drawn from a uniform distribution on $[-\omega_x, \omega_x]$ and $[-\omega_b, \omega_b]$, respectively. The input scaling ω_x , the bias scaling ω_b , the spectral radius ρ , and the leaking rate α are treated as hyper-parameters.

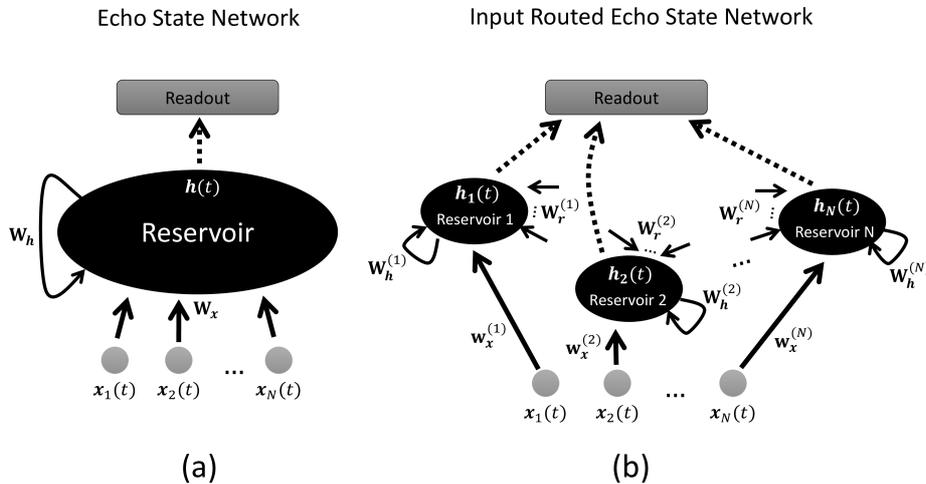


Fig. 1: Schematic representation of the standard ESN and of the introduced IR-ESN architectures. The dotted connections are the only ones that are trained. (a): All input components together drive the dynamics of a single reservoir. (b): Each component of the input signal is routed to a dedicated sub-reservoir.

We introduce a multi-reservoir architecture, called Input Routed Echo State

Network (IR-ESN), in which each dimension of the input signal is routed to a dedicated dynamical module. The dynamics of each sub-reservoir are driven by those of a one-dimensional input signal only, with additional incoming connections from the other sub-reservoirs. The architecture of IR-ESN is illustrated in Fig 1(b). The system state is split into N groups, each one describing the dynamical evolution of one sub-reservoir in the architecture. Using the superscript (i) to refer to the i -th sub-reservoir, the state transition equation of IR-ESN can be formulated as follows:

$$\begin{aligned} & \text{for } i = 1, \dots, N \\ \mathbf{h}^{(i)}(t) = & (1 - \alpha) \mathbf{h}^{(i)}(t - 1) + \alpha \tanh(\mathbf{W}_{\mathbf{h}}^{(i)} \mathbf{h}^{(i)}(t - 1) + \\ & \sum_{j \neq i} \mathbf{W}_r^{(i)} \mathbf{h}^{(j)}(t - 1) + \mathbf{w}_x^{(i)} \mathbf{x}_i(t) + \mathbf{b}^{(i)}), \end{aligned} \quad (2)$$

where $\mathbf{w}_x^{(i)} \in \mathbb{R}^{H^{(i)}}$ is the input weight vector, and $\mathbf{W}_r^{(i)} \in \mathbb{R}^{H^{(i)} \times H^{(j)}}$ is the inter-reservoir weight matrix that modulates the connections from all sub-reservoirs $j \neq i$ to sub-reservoir i . The readout is fed by the concatenation of all the sub-reservoir states, i.e., by $[\mathbf{h}^{(1)}; \dots; \mathbf{h}^{(N)}]$. The parameters in eq. 2 are initialized similarly to the standard ESN case, i.e. by controlling, for each sub-reservoir i , the spectral radius $\rho^{(i)} = \rho(\mathbf{W}_{\mathbf{h}}^{(i)})$, the input scaling $\omega_x^{(i)}$ (for the weights in $\mathbf{w}_x^{(i)}$), and the bias scaling $\omega_b^{(i)}$. In addition, the values in $\mathbf{W}_r^{(i)}$ are drawn from a uniform distribution on $[-\omega_r^{(i)}, \omega_r^{(i)}]$, where $\omega_r^{(i)}$ is an inter-reservoir scaling hyper-parameter, which modulates the strength of the inter-reservoir connections from the other sub-reservoirs. Differently from standard ESN, the input routed reservoir system in IR-ESN has the possibility to develop a temporal representation that can better focus on the dynamics of each individual component of the driving signal, while the dynamical interplay is modeled at the reservoir level by the inter-reservoir connections. The model also adds flexibility to the design, allowing different sub-reservoirs to have different scaling coefficients, hence to have a potentially different memory length and non-linear response behavior in correspondence to each input component¹. Moreover, at training stage, the readout has the possibility to modulate the contribution from each sub-reservoir, hence giving different weight to the dynamics influenced by different input dimensions in a task-dependent way.

In reporting our experimental analysis, we consider two settings of the proposed model. One, simply referred to as IR-ESN, consists in having different hyper-parameter values for each sub-reservoir. In the second, named IR-ESN (simple), we use the same values of the hyper-parameters in each sub-reservoir, i.e., $\rho^{(1)} = \dots = \rho^{(N)}$, $\omega_x^{(1)} = \dots = \omega_x^{(N)}$, $\omega_b^{(1)} = \dots = \omega_b^{(N)}$, $\omega_r^{(1)} = \dots = \omega_r^{(N)}$.

3 Experiments

Synthetic experiment. We first assessed the performance of IR-ESNs on a synthetic task, specifically designed to test the capability of the model to develop

¹The IR-ESN model can, in general, allow for different leaking rates for different sub-reservoirs. In this paper, for simplicity, we used a single leaking value α for all of them.

useful dynamics in presence of disturbing input of increasing dimensionality. As a reference task, we used the Cylinder-Bell-Funnel (CBF) [6], which is a sequence classification task on a one-dimensional input. Input time-series contain data drawn from standard normal noise, with the addition of a sub-sequence that has one of three possible offsets of a specific shape (cylinder, bell or funnel), which identifies the target class. The dataset contains 930 sequences in total, and we used its original splitting for training and test. We created variants of the CBF task by progressively adding a number of d additional input dimensions that contain standard normal noise, with $d = 1, \dots, 4$. The resulting dimension of the input signal is $N = d + 1$. For increasing d , the model is progressively challenged to develop dynamics that allow it to get rid of the increasing noise, and correctly classify the signals according to the only relevant input dimension.

We considered IR-ESN and IR-ESN (simple) with $H = 200$ reservoir neurons in total, divided into N sub-reservoirs of the same size, i.e. with H/N neurons each. When N is not an integer divisor of H , we reduced the total number of neurons by a sufficient amount to have sub-reservoirs of the same size. We explored values of the leaking rate $\alpha \in \{0.01, 0.1, 1\}$, of the spectral radius $\rho \in \{0.1, 0.3, 0.5, 0.7, 0.9, 1.1\}$, and of input scaling ω_x , bias scaling ω_b , and inter-reservoir scaling ω_r in $\{0, 0.001, 0.01, 0.1, 1\}$. For comparison, we ran experiments with standard ESNs using the same settings and the same ranges for the applicable hyper-parameters, arranging all recurrent units in one larger reservoir layer. For every model, the readout was implemented as a dense output layer with the same size as the number of target classes, and applied only to the last state computed for each input sequence. It was trained using Adam with learning rate 0.001 for a maximum number of 500 epochs, using early stopping with patience 10. We selected the values of the hyper-parameters independently for each model on a validation set drawn by stratified splitting (67/33) of the available training sequences, using Bayesian search with 300 trials. After model selection, we instantiated 5 networks with the selected configuration, training them on the training set and assessing their performance on the test set. The reported performance for each model is achieved by taking the average, and calculating std, of the results of the 5 repetitions.

Fig. 2 shows the test accuracy achieved by the considered models at the increase of the number of noisy input dimensions d . Results clearly show that IR-ESNs models significantly exceed the level of accuracy achievable by ESNs for all the cases explored. While the performance of ESNs degrades rapidly as the number of noisy dimensions increases, IR-ESNs are able to preserve a high level accuracy, showing the effectiveness of the proposed input-routed architecture in enabling reservoir dynamics representing the relevant input signal without incurring catastrophic interference from the others. Moreover, the additional flexibility in the hyper-parameters tuning granted in the full IR-ESN setting led to higher accuracy than the IR-ESN (simple) case.

Time-series classification benchmarks. The potential of the proposed architecture was then tested on a number of real-world classification tasks on multi-dimensional time-series. We considered 6 datasets from the UEA and UCR time-

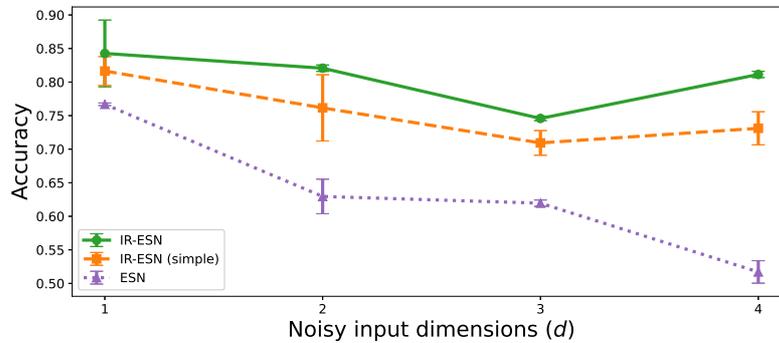


Fig. 2: Test set accuracy on the synthetic task derived from CBF, for increasing number of noisy input dimensions.

series classification repository [6], namely BasicMotions, CharacterTrajectories, Epilepsy, Libras, PhonemeSpectra, and UWaveGestureLibrary. They cover diverse applications in the area of human monitoring and activity recognition, from diverse signal sources including accelerometers, motion sensors, signals extracted from cameras, as well as frequency bands from audio recordings. An overview of the datasets properties is given in Table 1.

Name	# Seq.	Max L.	# Input dim.	# Classes
BasicMotions	80	100	6	4
CharacterTrajectories	2858	182	3	20
Epilepsy	275	206	3	4
Libras	360	45	2	15
PhonemeSpectra	6668	217	11	39
UWaveGestureLibrary	4479	315	3	8

Table 1: Summary of the datasets used, including the total number of sequences (# Seq), the maximum length of a sequence (Max L.), the number of input dimensions (# Input dim.), and the number of target classes (# Classes).

We ran experiments following the same settings used for the synthetic task, and we report the achieved test accuracy in Table 2. The results clearly highlight the competitiveness of the proposed approach, with IR-ESNs variants generally outperforming ESNs, sometimes even by a large margin. In particular, IR-ESN with variable reservoir hyper-parameters is the best performing method in most cases. The higher accuracy is paid, though, by a more involving hyper-parameters search in a higher-dimensional space. In some cases, this could limit the resulting final performance. And in fact, the two cases in which IR-ESN is not the best performing model are those featured by the higher input dimensionality, where the IR-ESN (simple) setting could reach a slightly better result.

Dataset	ESN	IR-ESN _(simple)	IR-ESN
BasicMotions	84.00 ± 2.55	100.00 ± 0.00	99.00 ± 1.22
CharacterTrajectories	95.31 ± 0.24	96.81 ± 0.51	97.31 ± 0.18
Epilepsy	86.96 ± 1.30	86.96 ± 1.65	88.84 ± 1.08
Libras	40.44 ± 1.94	35.44 ± 1.38	51.67 ± 2.17
PhonemeSpectra	7.00 ± 0.60	8.77 ± 0.46	8.73 ± 0.46
UWaveGestureLibrary	74.13 ± 1.07	70.50 ± 0.81	78.75 ± 1.33

Table 2: Test set accuracy achieved on the time-series classification benchmarks. Best results for each dataset are highlighted in bold.

4 Conclusions

In this paper, we have introduced a novel architecture for Reservoir Computing in the presence of multi-dimensional input time-series. The approach revolves around the idea of routing the different input dimensions towards different sub-reservoirs. These can develop dynamics that better focus on the individual components of the driving signal. The model also includes inter-reservoir connections whose strength can be modulated to accommodate the dynamical interactions that are needed for the specific task at hand. Through experiments on a synthetic task, we have shown that the proposed approach is significantly more robust than standard RC to disturbing input signals. The application potential of our proposal has then been confirmed by experiments on real-world tasks for multi-dimensional time-series classification.

Besides the evident advantages demonstrated already in the current IR-ESN form in this paper, our investigation indicated a potential limitation to scale to very high dimensional inputs, for which the hyper-parameter space would be impractical to explore. Nonetheless, beyond the methodology used in this paper, a general outcome of our study is that separating the role of the input signals in the model’s representations can be advantageous, especially when some of them might be noisy or less significant to the task. In the future, we intend to investigate this idea outside the RC framework, using learning as a viable way to configure the input routed recurrent architecture.

References

- [1] K. Nakajima and I. Fischer. *Reservoir Computing*. Springer, 2021.
- [2] A. Goudarzi and C. Teuscher. Reservoir computing: Quo vadis? In *Proceedings of NanoCom*, pages 1–6, 2016.
- [3] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert. Optimization and applications of echo state networks with leaky-integrator neurons. *Neural networks*, 20(3):335–352, 2007.
- [4] I. B Yildiz, H. Jaeger, and S. J Kiebel. Re-visiting the echo state property. *Neural networks*, 35:1–9, 2012.
- [5] C. Gallicchio, A. Micheli, and L. Pedrelli. Fast spectral radius initialization for recurrent neural networks. In *Proceedings of INNSBDDL*, pages 380–390, 2019.
- [6] A. Bagnall, J. L. W. Vickers, and E. Keogh. The uea & ucr time series classification repository. www.timeseriesclassification.com.