ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

Randomized Recurrent Neural Networks

Claudio Gallicchio[1], Alessio Micheli[1], Peter Tiňo[2]

1 - Department of Computer Science, University of Pisa
Largo Bruno Pontecorvo 3 - 56127 Pisa, Italy
2 - School of Computer Science , The University of Birmingham
Edgbaston, Birmingham B15 2TT, United Kingdom

Abstract. Neural Networks (NNs) with random weights represent nowadays a topic of consolidated use in the Machine Learning research community. In this contribution we focus in particular on recurrent NN models, which in a randomized setting represent a case of particular interest per se, entailing a number of intriguing research challenges primarily related to the control of the developed dynamics for the learning purposes.

Framed in the Reservoir Computing paradigm, this paper aims at providing the basic elements and summarizing the recent advances on the topic of randomized recurrent NNs (divided in theoretical, architectural, deep learning and structured domain aspects), and introducing the papers accepted for the ESANN special session.

1 Introduction

The use of randomization in the design of Neural Networks (NNs) has become increasingly popular, mainly due to the ease of implementation, extreme efficiency of the training algorithms and the possibility of analyzing the NNs properties that are independent from learning. Randomization can enter NN design in various disguises, for example in the model construction and training (e.g. random setting of a subset of weights as in the case of random projection NNs, including Extreme Learning Machines and Random Vector Functional-Link Networks, and Convolutional Neural Networks with random weights), or in its functionality and regularization algorithms (e.g. inclusion of random noise in activation layers, drop-out techniques, etc.). Under a broader perspective, the analysis of randomized models naturally extends to a general Machine Learning (ML) context (e.g. random projections, random forest models and random search for hyper-parameters selection, as well as generative approaches by generative adversarial networks). Moreover, Learning in Structured Domains and Deep Learning represent ML research areas for which this type of analysis is highly beneficial.

It is recognized, also through debates in the ML community, the need to target evidences, both in terms of theory or empirical results, to support the discussion on the advantages and limitations/shortcomings of the randomized approach. In the following, we aim to contribute to such critical analysis summarizing some recent advances in randomized NNs, with a focus on the area of Recurrent Neural Networks (RNNs) and Reservoir Computing, where the randomization has a peculiar role that can be better understood thanks to the

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

perspective of dynamical system analysis, including stability (see next sections) and learnability [1].

Randomization can characterize all the instances under the NN and RNN umbrella. E.g., randomization can be in the level of the output units with stochastic networks (as in the case of Boltzmann machines), or in the way in which the weights of hidden layer's connections are fixed (in both feed-forward and recurrent cases). In this paper we explicitly target the case of RNNs with random weights in the recurrent connections.

2 RNNs with Random Weights

The use of random weights in NNs is a practice that roots back to the early works on perceptrons. Under a general perspective, randomized NNs are composed of an untrained hidden layer, used to non-linearly project the input patterns into a high dimensional feature space, and of a trained readout component that exploits the hidden layer's representation to compute the output [2, 3]. The striking advantage with respect to fully trained NNs is naturally given by the fact that only the readout part needs to undergo a training process, e.g. by means of direct methods such as pseudo-inversion or Tikhonov regularization. This idea is exploited in the class of feed-forward NNs in different forms, including Extreme Learning Machines and Random Vector Functional-link Networks.

In the case of RNNs the paradigm is extended to cope with input of temporal/sequential nature. In this case, the untrained hidden layer contains recurrent units that, stimulated by an external signal, equip the network with a state-based memory of previous inputs, used by the readout component to modulate the network's output. The operation of the recurrent hidden layer can be understood as that of a non-linear dynamical system whose state evolution obeys a law that depends also on the driving input signal (i.e., it is a non-autonomous system). Such a system can be described through differential equations (in the continuous-time case) or by means of iterated mappings (for discrete-time systems). In both cases, the crucial characterization is that the network's dynamics are parameterized by a set of weights that are randomly initialized and then are left untrained. Accordingly, one of the fundamental issues in randomized RNNs consists in ensuring the stability of the state behavior, which can be attained e.g. by imposing contractive properties to the developed network dynamics. One major example is given by the Echo State Property (ESP) [4, 5], defined in the context of Echo State Networks (ESNs) [6]. Essentially, the ESP says that the state evolution should asymptotically depend on the driving input signal, whereas the influence of initial conditions should progressively die out. Conditions for the ESP are traditionally given in terms of spectral properties of the recurrence matrix of the system, i.e. the matrix that collects the weights of the recurrent connections, thereby resulting in a simple and easy to implement initialization process. Analogous properties were also provided for Liquid State Machines [7], in the context of spiking neurons, and for Fractal Prediction Machines [8], in the context of iterative function systems. These approaches are

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

now collectively known in literature under the name of Reservoir Computing (RC) methods [9, 10], where the untrained recurrent hidden layer is commonly referred to as the *reservoir*. Noticeably, the characterization of reservoir dynamics is also strictly related to the Markovian architectural bias of recurrent models [11, 12].

3 Advances

Based on the background presented in previous Section 2, here we give a brief overview of the recent advances and trends in the research on randomized recurrent models, with a specific attention to RC approaches. The major identified topics are described in the following sub-sections, which respectively focus on studies of theoretical nature (Section 3.1), explorations of novel reservoir architectures (Section 3.2), criteria for assessing the goodness of randomized dynamics (Section 3.3), deep models (Section 3.4) and extensions to complex structured data processing (Section 3.5).

3.1 Theoretical Analysis

Much of the theoretical research in RC is aimed at a deeper comprehension of the characteristics of the state dynamics that are developed by the fixed untrained reservoir component. A very important topic in this regard is of course represented by the study of the stability conditions for the reservoir dynamics and the ESP. Important contribution to this research line are based on the study of the reservoir behavior under the perspective of dynamical system theory. In particular, while initial studies tended to focus on the algebraic properties of the recurrence reservoir matrix, as in the original formulation of conditions for the ESN [4] and in the tighter bound developed later on [13], more and more effort has been developed in recent years to understand the conditions under which the reservoir develops stable dynamics *given the driving external input.* A first milestone in this regard is represented by the work in [5], which studied bifurcations of input driven reservoir dynamics and gave a novel condition for the ESP based on diagonal Schur stability of the recurrence matrix. Successively, framing the RC approach in the field of non-autonomous dynamical systems, the study in [14] for the first time derived a condition for the ESP that is directly linked to the characteristics of the input signal (under the assumption of ergodicity in the input generation process).

A further theoretical support is given by the frameworks of mean field theory, extended to the case of ESNs in [15], and of random matrix theory (see e.g. [16]). Insights from these areas can be extremely useful to characterize (and simplify the analysis of) the behavior of reservoirs in the limit of an increasingly large state dimension. Noticeable examples of recent works in this concern are given by [17], which derives a local operational version of the ESP given the external input based on controlling the largest Lyapunov exponent, by [18], which provides a theoretical analysis of the asymptotic performance of linear ESNs, and by the

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

work in [19, 20], in which asymptotic properties of the Fisher memory in linear ESNs are analyzed.

3.2 Architectural Studies

The theoretical studies mentioned in Section 3.1 are of huge relevance in the quest for a deeper understanding and of the behavior of input driven randomized dynamical neural models, with a potentially tremendous impact on applications to real-world problems. However, under the perspective of RC networks implementation and usage, the initially proposed methodology for reservoir initialization based on controlling the spectral radius of the recurrence matrix is still unparalleled, and major theoretical breakthroughs resulting in comparably simple network initialization recipes are still demanded. Under a more practical perspective, a constantly flourishing research line consists in investigating novel patterns of connectivity among the recurrent units in reservoir systems, analyzing the involved algebraic properties of the resulting recurrent reservoir weight matrix. The general aim in this case is to constrain the reservoir randomized construction towards conditions that can result in optimized performance. A major example is given by the study of reservoirs with orthogonal recurrent matrices [21], which also represent an instance of critical ESN systems [22]. A simple strategy to build orthogonal ESNs, as proposed in [22], is based on the use of permutation matrices to determine the pattern of connectivity among the reservoir units. As experimentally shown in [23], such a simple strategy has a beneficial effect both on the memory skills and on the predictive performance of the resulting networks, as measured on standard time-series benchmarks in the RC area. Recently, learning strategies to the orthogonalization of reservoirs have been explored in [24, 25], showing excellent performance in terms of memory ability. Another relevant trend in shaping the recurrent connectivity is related to the design of reservoirs organized in (possibly interconnected) groups, or clusters, of recurrent units, as investigated e.g. in [26, 27, 28]. A further perspective to the analysis of reservoir connectivity pattern was provided in [12] in relation to the Markovian characterization of ESN dynamics. In particular, the Markovian analysis of RC models provides a clear example of the effect on the identification of easy or hard tasks for randomized RNN subject to a stability assumption/constraint, which hence characterizes advantages and possible limitations of such approaches.

Other studies approached the analysis of reservoirs topology with the aim of simplifying the architecture and the general design of standard ESNs without affecting the performance in applications. Among the others, reservoirs whose recurrent units are organized in delay lines or in cyclic-based structures [29, 30, 31, 32] are prominent representatives. The importance of these studies is recently further increased by the fact that reservoirs with such a simplified pattern of connectivity are naturally amenable to physical implementations, e.g. in photonic realizations of the involved dynamics (see e.g. [33]).

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

3.3 Quality of Dynamics

Strictly linked to the exploration of novel reservoir design strategies is the research on measures of reservoir quality. Examples in this regard are given by the kernel quality and generalization ability [34, 35], by the average state entropy [36] and by the entropy of recurrent units' output distribution, the maximization of which led to the well known intrinsic plasticity adaptation rules [37, 38]. Another well known measure of reservoir quality is the short-term memory capacity introduced in [39], which quantifies the ability to reconstruct delayed versions of the input from the network's state. While this measure can be maximized by resorting to a linear activation function [39], the use of non-linear dynamics is important in order to achieve good performances in many complex applications, leading to what is known as the *memory versus non-linearity* trade-off [40, 41].

A much investigated condition of recurrent networks behavior is the so called edge of stability, or *edge of chaos* [42, 43, 44], where the internal state of the network crosses the border of transition between stable and unstable dynamical regimes. Several works in literature reported evidences that in proximity of such transition, RNNs tend to develop "high quality" dynamics that provide rich representation of their input history [42, 45, 43], resulting in maximized networks performance in terms of short-term memory [46, 24] and predictive performance on several tasks of different nature (see, e.g. [35, 10]). Adopted strategies to detect the edge of stability often try to account for the external driving input signal by analyzing the spectrum of Lyapunov exponents of reservoir states (see e.g. [46, 47]), whereas more recently devised approaches are based on the inspection of recurrence plots [48] and Fisher information maximization [49].

3.4 Deep Models

The investigation on deep RNN models is a research topic that is raising an increasing research attention in the neural networks community [50]. Deep RNNs proved able to learn temporal representations at different time-scales, achieving very good results in tasks related, e.g., to document and speech processing [51, 52]. However, training of RNNs with many non-linear recurrent layers amplifies the difficulties already observed for shallow models [53], and often requires ad-hoc strategies to be effective [51], as well as high performance computing support to achieve affordable training times in practice. It is therefore easy to understand that the possibility of developing efficient yet effective deep recurrent models represents a topic of extreme interest.

Randomized RNNs can again be a key factor to this aim. In particular, in the context of RC, different forms of hierarchical architectures have been explored, leading to performance improvements in benchmark as well as real-world tasks [9, 54, 55]. Recently, the RC framework has been explicitly extended towards deep architectures with the introduction of the DeepESN model [56, 57]. As in the case of standard (shallow) ESN architectures, a DeepESN is composed of a dynamical untrained part and of a feed-forward trained readout, with the fundamental difference that the dynamical component is in this case a stacked composition of *multiple recurrent reservoirs*. Inheriting, and even amplifying the efficiency of the RC approach, studies on DeepESNs allow on the one hand to investigate the intrinsic role played by layering in RNNs (even in the absence of training of the recurrent connections), and on the other hand enable the development of an extremely efficient approach for designing deep neural models in the temporal domain.

The major outcomes of the studies on DeepESNs conducted so far (see [58] for an updated view on this research topic) showed that stacking layers of recurrent units drives the network's dynamics to develop *multiple time-scales* [56] and *multi-frequency* [59] temporal representations of the input signal, within a randomized approach for the recurrent part. A hierarchical architectural construction of deep RNN models is reflected into an intrinsically structured state space organization. Noticeably, this effect was observed even in the case of (randomized) linear recurrent units, thereby opening intriguing questions on the true essence of deep learning for sequential/temporal processing [59].

On the theoretical side, the generalization of the ESP for the case of deep RC models [60] pointed out that higher layers in deep RNN architectures are naturally driven towards less contractive dynamics, even in simple settings of the hyper-parameters. Moreover, studies on the stability of deep recurrent models in presence of driving external input, conducted through the analysis of the maximum local Lyapunov exponent in [61, 62], showed that layering in RNN has a noticeable effect on the quality of network dynamics. If the same amount of recurrent units is organized in a hierarchy of levels, the resulting network's dynamical regime is pushed closer to the edge of stability (see Section 3.3). A straightforward implication of this phenomenon is that layering results as a convenient strategy for RNN architectural construction, e.g. whenever the task to be approached is known to require much in terms of memory [61]. The advantage of DeepESNs over standard shallow ESNs in terms of longer short-term memory has been investigated in [56, 60, 61], and more recently on a layer-wise basis in [63]. Though recently proposed, the DeepESN approach already proved effective in applications, showing competitive state-of-the-art results on both benchmarks and real-world tasks of heterogeneous nature [59, 64, 65]. Overall, investigations on deep RC models laid the foundations of a framework for efficient modeling of deep neural networks for temporal data processing, in which novel advances under the perspectives of theoretical analysis, architectural setup and applications can further contribute.

3.5 Randomized Networks for Structures

Many real-world problems involve data that can be naturally represented in form of structures, such as trees or graphs. Recursive Neural Networks (RecNNs) [66, 67] generalize the applicability of RNN models to directly deal with such structures. Specifically, extending the idea of temporal unfolding of RNNs, in the case of RecNNs the same recurrent architecture is unfolded over the nodes of the input structure. The approach, which has interesting theoretical properties (see e.g. [68]), has proved successful in several challenging real-world problems

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

in several application domains, including, among the others, cheminformatics, document and natural language processing. While opening a wide range of application possibilities, the extension of the input domain from vectors and sequences to highly complex data entails an explosion of training costs that makes a randomized approach a desirable option also in this field.

A first step in this direction was taken by the introduction of the Tree ESN [69], which generalizes the RC methodology to the case of tree data processing. Another aspect to take into account is related to the case of undirected or cyclic structures (as occurring, e.g., in general graph processing). In this case, the process of state computation presents cyclic dependencies and can be described in terms of the evolution of a dynamical system, for which stability conditions play of course a fundamental role. Leveraging on such studies, the Graph ESN [70] was proposed as an efficient approach for learning in graph structured data. Though randomized network approaches for structured data already proved effective in real-world applications, the feeling is that this field of research still has not expressed its full potential power, and further relevant improvements and achievements are still expected.

4 Special Session Contributions

As outlined in the previous Section 3, the study of RNN with randomized recurrent dynamics (and RC in particular) constitutes a rich and exciting research area, whose vitality is further testified by the works presented in the ESANN special session on randomized neural networks. In what follows, the major contributions of the accepted papers are briefly recalled.

In [71], Bianchi et al. propose a novel RC architecture, that combines a bi-directional reservoir with a deep feed-forward readout. Learning is facilitated by a preliminary step of dimensionality reduction, performed through principal component analysis of reservoir states. Experimental assessment on several benchmarks, as well as on a real-world task in the medical area, show that the approach achieves competitive predictive performance with respect to state-of-the-art networks, still preserving the training efficiency advantage typical of RC.

An application of advanced RC methodology in the area of financial forecasting is presented in [72], in which Rodan et al. explore the use of deterministically constructed delay line reservoirs in conjunction with readout components of different nature. Interestingly, the authors exploit the simplified architectural setup for efficient adaptation of reservoir parameters. As it can be read in the paper, experimental results on a task of bankruptcy prediction show the competitiveness of the RC approach in a real-world complex scenario.

In [73], Dashdamirov et al. present an application of ESNs to a real-world problem related to the estimation of the level of human concentration. Specifically, the authors deal with a binary classification task in which streams of EEG signals are labeled as corresponding to high mental focus or not. Although somewhat preliminary, the experimental assessment reported in the paper already shows encouraging results and a good accuracy in a basic ESN setup.

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

In [74], Burianek and Basterrech propose to assess the quality of reservoirs in ESNs by using a methodology based on Sammon projection. Through experiments on some benchmark datasets on time-series, the authors observe a correlation between the predictive accuracy and the Sammon's energy evaluated between the reservoir states and a time-windowed version of the input streams.

References

[1] B. Hammer and P. Tiňo. Recurrent neural networks with small weights implement definite memory machines. *Neural Computation*, 15(8):1897–1929, 2003.

[2] C. Gallicchio, J.D. Martin-Guerrero, A. Micheli, and E. Soria-Olivas. Randomized machine learning approaches: Recent developments and challenges. In *Proceedings of ESANN*, pages 77–86. i6doc.com, 2017.

[3] S. Scardapane and D. Wang. Randomness in neural networks: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(2), 2017.

[4] H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148(34):13, 2001.

[5] I.B. Yildiz, H. Jaeger, and S.J. Kiebel. Re-visiting the echo state property. *Neural networks*, 35:1–9, 2012.

[6] H. Jaeger and H. Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *science*, 304(5667):78–80, 2004.

[7] W. Maass, T. Natschläger, and H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, 14(11):2531–2560, 2002.

[8] P. Tiňo and G. Dorffner. Predicting the future of discrete sequences from fractal representations of the past. *Machine Learning*, 45(2):187–217, 2001.

[9] M. Lukoševičius and H. Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.

[10] D. Verstraeten, B. Schrauwen, M. d'Haene, and D. Stroobandt. An experimental unification of reservoir computing methods. *Neural networks*, 20(3):391–403, 2007.

[11] P. Tiňo, M. Cernansky, and L. Benuskova. Markovian architectural bias of recurrent neural networks. *IEEE Transactions on Neural Networks*, 15(1):6–15, 2004.

[12] C. Gallicchio and A. Micheli. Architectural and markovian factors of echo state networks. *Neural Networks*, 24(5):440–456, 2011.

[13] M. Buehner and P. Young. A tighter bound for the echo state property. *IEEE Transactions on Neural Networks*, 17(3):820–824, 2006.

[14] G. Manjunath and H. Jaeger. Echo state property linked to an input: Exploring a fundamental characteristic of recurrent neural networks. *Neural computation*, 25(3):671–696, 2013.

[15] M. Massar and S. Massar. Mean-field theory of echo state networks. *Physical Review E*, 87(4):042809, 2013.

[16] T. Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

[17] G. Wainrib and M.N. Galtier. A local echo state property through the largest lyapunov exponent. *Neural Networks*, 76:39–45, 2016.

[18] R. Couillet, G. Wainrib, H. Sevi, and H.T. Ali. The asymptotic performance of linear echo state neural networks. *Journal of Machine Learning Research*, 17(178):1–35, 2016.

[19] P. Tiňo. Asymptotic fisher memory of randomized linear symmetric echo state networks. *Neurocomputing*, 2018.

[20] P. Tiňo. Fisher memory of linear wigner echo state networks. In i6doc. com, editor, *Proceedings of ESANN*, pages 87–92, 2017.

[21] O.L. White, D.D. Lee, and H. Sompolinsky. Short-term memory in orthogonal neural networks. *Physical review letters*, 92(14):148102, 2004.

[22] M.A. Hajnal and A. Lőrincz. Critical echo state networks. In *International Conference on Artificial Neural Networks*, pages 658–667. Springer, 2006.

[23] J. Boedecker, O. Obst, N.M. Mayer, and M. Asada. Studies on reservoir initialization and dynamics shaping in echo state networks. In d side, editor, *Proceedings of ESANN*, pages 227–232, 2017.

[24] I. Farkaš, R. Bosák, and P. Gergel'. Computational analysis of memory capacity in echo state networks. *Neural Networks*, 83:109–120, 2016.

[25] I. Farkaš and P. Gergel'. Maximizing memory capacity of echo state networks with orthogonalized reservoirs. In *Proceedings of IJCNN*, pages 2437–2442. IEEE, 2017.

[26] Y. Xue, L. Yang, and S. Haykin. Decoupled echo state networks with lateral inhibition. *Neural Networks*, 20(3):365–376, 2007.

[27] J. Qiao, F. Li, H. Han, and W. Li. Growing echo-state network with multiple subreservoirs. *IEEE transactions on neural networks and learning systems*, 28(2):391–404, 2017.

[28] F.M. Bianchi, M. Kampffmeyer, E. Maiorino, and Robert R. Jenssen. Temporal overdrive recurrent neural network. In *Proceedings of IJCNN*, pages 4275–4282. IEEE, 2017.

[29] M. Čerňanský and P. Tiňo. Predictive modeling with echo state networks. In *International Conference on Artificial Neural Networks*, pages 778–787. Springer, 2008.

[30] A. Rodan and P. Tiňo. Simple deterministically constructed cycle reservoirs with regular jumps. *Neural computation*, 24(7):1822–1852, 2012.

[31] A. Rodan and P. Tiňo. Minimum complexity echo state network. *IEEE transactions on neural networks*, 22(1):131–144, 2011.

[32] T. Strauss, W. Wustlich, and R. Labahn. Design strategies for weight matrices of echo state networks. *Neural computation*, 24(12):3246–3276, 2012.

[33] G. Van der Sande, D. Brunner, and M.C. Soriano. Advances in photonic reservoir computing. *Nanophotonics*, 6(3):561–576, 2017.

[34] W. Maass, R.A. Legenstein, and N. Bertschinger. Methods for estimating the computational power and generalization capability of neural microcircuits. In *Advances in neural information processing systems*, pages 865–872, 2005.

[35] B. Schrauwen, L. Büsing, and R.A. Legenstein. On computational power and the order-chaos phase transition in reservoir computing. In *Advances in Neural Information Processing Systems*, pages 1425–1432, 2009.

[36] M.C. Ozturk, D. Xu, and J. Príncipe. Analysis and design of echo state networks. *Neural computation*, 19(1):111–138, 2007.

[37] J. Triesch. Synergies between intrinsic and synaptic plasticity in individual model neurons. In *Advances in neural information processing systems*, pages 1417–1424, 2005.

[38] B. Schrauwen, M. Wardermann, D. Verstraeten, J.J. Steil, and D. Stroobandt. Improving reservoirs using intrinsic plasticity. *Neurocomputing*, 71(7-9):1159–1171, 2008.

[39] H. Jaeger. *Short term memory in echo state networks*, volume 5. GMD-Forschungszentrum Informationstechnik, 2001.

[40] D. Verstraeten, J. Dambre, X. Dutoit, and B. Schrauwen. Memory versus non-linearity in reservoirs. In *Proceedings of IJCNN)*, pages 1–8. IEEE, 2010.

[41] M. Inubushi and K. Yoshimura. Reservoir computing beyond memory-nonlinearity trade-off. *Scientific Reports*, 7(1):10199, 2017.

[42] N. Bertschinger and T. Natschläger. Real-time computation at the edge of chaos in recurrent neural networks. *Neural computation*, 16(7):1413–1436, 2004.

[43] R. Legenstein and W. Maass. What makes a dynamical system computationally powerful. *New directions in statistical signal processing: From systems to brain*, pages 127–154, 2007.

[44] T. Yamane, S. Takeda, D. Nakano, G. Tanaka, R. Nakane, S. Nakagawa, and A. Hirose. Dynamics of reservoir computing at the edge of stability. In *Neural Information Processing*, pages 205–212. Springer International Publishing, 2016.

[45] R. Legenstein and W. Maass. Edge of chaos and prediction of computational performance for neural circuit models. *Neural Networks*, 20(3):323–334, 2007.

[46] J. Boedecker, O. Obst, J.T. Lizier, N.M. Mayer, and M. Asada. Information processing in echo state networks at the edge of chaos. *Theory in Biosciences*, 131(3):205–213, 2012.

[47] D. Verstraeten and B. Schrauwen. On the quantification of dynamics in reservoir computing. In *Artificial Neural Networks – ICANN 2009*, pages 985–994. Springer Berlin Heidelberg, 2009.

[48] F. M. Bianchi, L. Livi, and C. Alippi. Investigating echo-state networks dynamics by means of recurrence analysis. *IEEE Transactions on Neural Networks and Learning Systems*, 29(2):427–439, 2018.

[49] L. Livi, F. M. Bianchi, and C. Alippi. Determination of the edge of criticality in echo state networks through fisher information maximization. *IEEE Transactions on Neural Networks and Learning Systems*, PP(99):1–12, 2018.

[50] P. Angelov and A. Sperduti. Challenges in deep learning. In *Proceedings of ESANN*, pages 489–495. i6doc.com, 2016.

[51] M. Hermans and B. Schrauwen. Training and analysing deep recurrent neural networks. In *NIPS*, pages 190–198, 2013.

[52] A. Graves, A.-R. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. IEEE, 2013.

[53] R. Pascanu, T. Mikolov, and Y. Bengio. On the difficulty of training recurrent neural networks. In *Proceedings of ICML*, pages 1310–1318, 2013.

[54] F. Triefenbach, A. Jalalvand, K. Demuynck, and J.-P. Martens. Acoustic modeling with hierarchical reservoirs. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(11):2439–2450, 2013.

[55] F. Triefenbach, A. Jalalvand, B. Schrauwen, and J.-P. Martens. Phoneme recognition with large hierarchical reservoirs. In *Advances in neural information processing systems*, pages 2307–2315, 2010.

[56] C. Gallicchio, A. Micheli, and L. Pedrelli. Deep reservoir computing: a critical experimental analysis. *Neurocomputing*, 268:87–99, 2017.

[57] C. Gallicchio and A. Micheli. Deep reservoir computing: A critical analysis. In *Proceedings of ESANN*, pages 497–502. i6doc.com, 2016.

[58] C. Gallicchio and A. Micheli. Deep echo state network (DeepESN): A brief survey. *arXiv preprint arXiv:1712.04323*, 2017.

[59] C. Gallicchio, A. Micheli, and L. Pedrelli. Hierarchical temporal representation in linear reservoir computing. In *Proceedings of WIRN*, 2017. arXiv preprint arXiv:1705.05782.

[60] C. Gallicchio and A. Micheli. Echo state property of deep reservoir computing networks. *Cognitive Computation*, 9:337–350, 2017.

[61] C. Gallicchio, A. Micheli, and L. Silvestri. Local lyapunov exponents of deep echo state networks. *Neurocomputing*, 2017.

[62] C. Gallicchio, A. Micheli, and L. Silvestri. Local Lyapunov Exponents of Deep RNN. In *Proceedings of ESANN*, pages 559–564. i6doc.com, 2017.

[63] C. Gallicchio. Short-term memory of deep rnn. In *Proceedings of ESANN*, 2018.

[64] C. Gallicchio and A. Micheli. Experimental analysis of deep echo state networks for ambient assisted living. In *Proceedings of AI*AAL 2017, co-located with the AI*IA 2017*, 2017.

[65] C. Gallicchio, A. Micheli, and L.Pedrelli. Deep echo state networks for diagnosis of parkinson's disease. In *Proceedings of ESANN*, 2018.

[66] A. Sperduti and A. Starita. Supervised neural networks for the classification of structures. *IEEE Transactions on Neural Networks*, 8(3):714–735, 1997.

[67] P. Frasconi, M. Gori, and A. Sperduti. A general framework for adaptive processing of data structures. *IEEE transactions on Neural Networks*, 9(5):768–786, 1998.

[68] B. Hammer, A. Micheli, and A. Sperduti. Adaptive contextual processing of structured data by recursive neural networks: A survey of computational properties. In *Perspectives of Neural-Symbolic Integration*, pages 67–94. Springer, 2007.

[69] C. Gallicchio and A. Micheli. Tree echo state networks. *Neurocomputing*, 101:319–337, 2013.

[70] C. Gallicchio and A. Micheli. Graph echo state networks. In *Proceedings of IJCNN*, pages 1–8. IEEE, 2010.

[71] F.M. Bianchi, S. Scardapane, S. Lokse, and R. Jenssen. Bidirectional deep-readout echo state networks. In *Proceedings of ESANN*, 2018.

[72] F.M. Bianchi, P.A. Castillo, H. Faris, Ala' M. Al-Zoubi, A.M. Mora, and H. Jawazneh. Forecasting business failure in highly imbalanced distribution based on delay line reservoir. In *Proceedings of ESANN*, 2018.

[73] H. Dashdamirov and S. Basterrech. Estimation of human concentration using echo state networks. In *Proceedings of ESANN*, 2018.

[74] T. Burianek and S. Basterrech. Estimation of human concentration using echo state networks. In *Proceedings of ESANN*, 2018.