

In addition to provide insights on such general issues, this paper would contribute to show a practical way to efficiently approach the design of learning models in the scenario of deep RNN, extending the set of tools available to the users for complex time-series tasks. Indeed, the first empirical results provided in this paper seem to indicate that some classes of models are sometimes uncritically adopted, i.e. despite their cost, guided by the natural popularity due to their software availability (GRU, LSTM). The same diffusion of software tools deserve more effort on the side of the other models (DeepESN class), although the first instances are already available².

References

- [1] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [2] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [3] M. Lukoševičius and H. Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.
- [4] C. Gallicchio, A. Micheli, and L. Pedrelli. Deep reservoir computing: a critical experimental analysis. *Neurocomputing*, 268:87–99, 2017.
- [5] C. Gallicchio, A. Micheli, and L. Pedrelli. Hierarchical Temporal Representation in Linear Reservoir Computing. In *Neural Advances in Processing Nonlinear Dynamic Signals*, pages 119–129, Cham, 2019. Springer International Publishing. WIRN 2017.
- [6] C. Gallicchio, A. Micheli, and L. Pedrelli. Design of deep echo state networks. *Neural Networks*, 108:33 – 47, 2018.
- [7] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent. Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription. *arXiv preprint arXiv:1206.6392*, 2012.
- [8] Y. Bengio, N. Boulanger-Lewandowski, and R. Pascanu. Advances in optimizing recurrent networks. In *ICASSP 2013*, pages 8624–8628. IEEE, 2013.
- [9] H. Jaeger and H. Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004.
- [10] C. Gallicchio and A. Micheli. Echo State Property of Deep Reservoir Computing Networks. *Cognitive Computation*, 9(3):337–350, 2017.
- [11] D. Kinga and J. B. Adam. A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, volume 5, 2015.

²DeepESN implementations are made publicly available for download both in MATLAB (see <https://it.mathworks.com/matlabcentral/fileexchange/69402-deepesn>) and in Python (see <https://github.com/lucapedrelli/DeepESN>).