Recurrent Neural Networks Models for Developmental Language Acquisition: Reservoirs Outperform LSTMs
Xavier Hinaut, Alexandre Varengien

To cite this version:
Xavier Hinaut, Alexandre Varengien. Recurrent Neural Networks Models for Developmental Language Acquisition: Reservoirs Outperform LSTMs. SNL 2020 - 12th Annual Meeting of the Society for the Neurobiology of Language, Oct 2020, Virtual Edition, Canada. hal-03146558

HAL Id: hal-03146558
https://inria.hal.science/hal-03146558
Submitted on 19 Feb 2021
Recurrent Neural Networks Models for Developmental Language Acquisition: Reservoirs Outperform LSTMs

Xavier Hinaut\(^1,2,3\)\(^*,\) Alexandre Variengien\(^1,2,3,4\)

\(^1\)Inria, Bordeaux, France. \(^2\)LaBRI, UMR 5800, Talence, France. \(^3\)Neurodegenerative Diseases Institute, UMR 5293, Bordeaux, France. \(^4\)Ecole Normal Supérieure de Lyon, France. \(*\)xavier.hinaut@inria.fr

INTRODUCTION

We previously developed cortico-striatal models for sentence comprehension (Hinaut & Dominey 2013) and sentence production (Hinaut et al. 2015). The sentence comprehension model is based on the reservoir computing principle (Jaeger & Haas 2004): a random recurrent neural network (a reservoir) provides a rich recombination of sequential word inputs (e.g. a piece of prefrontal cortex), and an output layer (e.g. striatum) learns to "read-out" the roles of words in the sentence via the internal recurrent dynamics. The model has several interesting properties, like the ability to predict the semantic roles of words during online processing. Additionally, we demonstrated its robustness to various corpus complexities, in different languages (Hinaut & Tanié 2019), and even its ability to work with bilingual inputs.

ABSTRACT

In this study, we propose to (1) use the model in a new task related to a developmental language acquisition (i.e. Cross-Situational Learning) (Juven & Hinaut 2020; Diniz & Hinaut 2020), (2) provide a qualitative comparison with one of the best performing neural networks for sequential tasks (a LSTM), and (3) provide a qualitative analysis on the way reservoirs and LSTMs solve the task. This new Cross-Situational Task is as follows: for a given sentence, the target output provided often contains more detailed features than what is available in the sentence. Thus, the models have not only to learn how to parse sentences to extract useful information, but also to statistically infer which word is associated with which feature. While reservoir units are modelled as leaky average firing rate neurons, LSTM units (Hochreiter & Schmidhuber, 1997) are engineered to gate information using a costly and biologically implausible learning algorithm (Back-Propagation Through Time).

RESULTS

We found that both models were able to successfully learn the task: the LSTM reached slightly better performance for the basic corpus, but the reservoir was able to significantly outperform LSTM on more challenging corpora with increasing vocabulary sizes (for a given set of hyperparameters). We analysed the hidden activations of internal units of both models. Despite the deep differences between both models (trained or fixed internal weights), we were able to uncover similar inner dynamics: the most useful units (with strongest weights to the output layer) seemed tuned to keep track of several specific words in the sentence. Because of its learning algorithm, it is predictable to see such behaviour in a LSTM but not in a reservoir; in fact, the LSTM contained more tuned-like units than the reservoir.

CONCLUSION

The qualitative and quantitative differences between LSTMs and reservoirs highlight the gap between classical Deep Learning approaches (based on back-propagation algorithm) and more plausible brain learning mechanisms. First, the reservoir is more efficient in terms of training time and cost (the LSTM needs several passes on the training data, while the reservoir uses only one). Secondly, only the reservoir model seems to scale to larger corpora without the need to adapt specifically the hyper-parameters of the model. Finally, the presence of more tuned units in the LSTM compared to the reservoir might be an explanation of why the LSTM outperforms the training data. More generally, a LSTM seems to have more limited generalization capabilities than a reservoir when available data is limited.