2016 Joint IEEE International Conference on Development and Learning
and Epigenetic Robotics (ICDL-EpiRob)
Cergy-Pontoise, Paris, France, Sept 19-22, 2016

Recurrent Neural Network for Syntax Learning with Flexible Predicates for Robotic Architectures

Xavier Hinaut[1,2], Johannes Twiefel[1], and Stefan Wermter[1]

Abstract— **We present a Recurrent Neural Network (RNN), namely an Echo State Network (ESN), that performs sentence comprehension and can be used for Human-Robot Interaction (HRI). The RNN is trained to map sentence structures to meanings (i.e. predicates). We have previously shown that this ESN is able to generalize to unknown sentence structures. Moreover, it is able to learn English, French or both at the same time. The are two novelties presented here: (1) the encapsulation of this RNN in a ROS module enables one to use it in a robotic architecture like the Nao humanoid robot, and (2) the flexibility of the predicates it can learn to produce (e.g. extracting adjectives) enables one to use the model to explore language acquisition in a developmental approach.**

I. INTRODUCTION

How do children learn language? In particular, how do they associate the structure of a sentence to its meaning? This question is linked to the more general issue: how does the brain associate sequences of symbols to internal symbolic or sub-symbolic representations? We propose a framework to understand how language is acquired based on a simple and generic neural architecture (Echo State Networks) [1] which is not hand-crafted for a particular task, but on the contrary can be used for a broad range of applications.

First, we present the general ESN architecture. Then, we present the reservoir sentence processing model and provide some examples of its flexibility. Finally, we show how to use it as a ROS module.

II. METHODS & RESULTS

A. Echo State Networks

The language module is based on an ESN [1] with leaky neurons: inputs are projected to a random recurrent layer and a linear output layer (called "read-out") is modified by learning (which can happen online). The units of the recurrent neural network have a *leak rate* (α) hyper-parameter which corresponds to the inverse of a time constant. These equations define the update of the ESN:

$$\mathbf{x}(t+1) = (1-\alpha)\mathbf{x}(t) + \alpha f(W^{in}\mathbf{u}(t+1) + W\mathbf{x}(t)) \quad (1)$$

$$\mathbf{y}(t) = W^{out}\mathbf{x}(t) \quad (2)$$

with $\mathbf{x}(t)$, $\mathbf{u}(t)$ and $\mathbf{y}(t)$ the reservoir (i.e. hidden) state, the input and the read-out (i.e. output) states respectively at time t, α the *leak rate*, W, W^{in} and W^{out} the reservoir, the input and the output matrices respectively and f the *tanh* activation function. After the collection of all reservoir states the following equation defines how the read-out (i.e. output) weights are trained:

$$W^{out} = Y^d[1; X]^+ \quad (3)$$

with Y^d the concatenation of the desired outputs, X the concatenation of the reservoir states (over all time steps for all train sentences) and M^+ the Moore-Penrose pseudo-inverse of matrix M. Hyper-parameters that can be used for this task are the following: *spectral radius*: 1, *input scaling*: 0.75, *leak rate*: 0.17, *number of reservoir units*: 100.

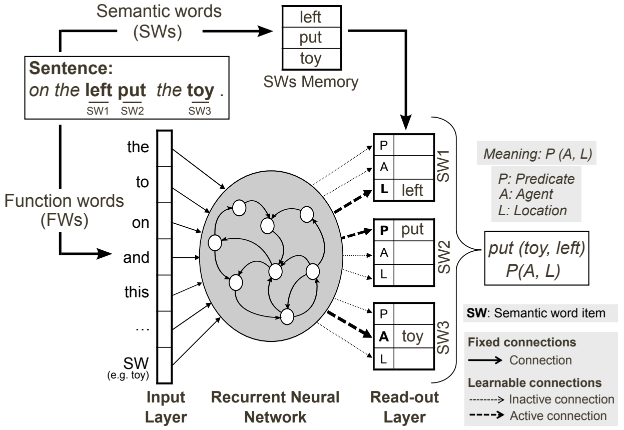

Fig. 1. Sentences are converted to a sentence structure by replacing semantic words by a SW marker. The ESN is given the sentence structure word by word. Each word activates a different input unit. During training, the connections to the readout layer are modified to learn the mapping between the sentence structure and the arguments of the predicates. When a sentence is tested, the most active units are bound with the SW kept in the SWs memory to form the resulting predicate. (Adapt. from [2].)

B. Reservoir Sentence Processing Model

The reservoir sentence processing model has been adapted from previous experiments on a neuro-inspired model for sentence comprehension using ESN [3] and its application to HRI [2]. The model learns the mapping of the semantic words (SW; e.g. nouns, verbs) of a sentence onto the different slots (the thematic roles: e.g. action, location) of a basic event structure (e.g. *action(object, location)*). As depicted in Fig. 1, the system processes a sentence as input and generates corresponding predicates. Before being fed to the

*This work was supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme: EchoRob project (PIEF-GA-2013-627156).

[1] X. Hinaut, J. Twiefel, and S. Wermter are with Knowledge Technology Group, Department Informatik, Universität Hamburg, 22527 Hamburg, Germany {surname}@informatik.uni-hamburg.de

[2] X. Hinaut is with Inria Bordeaux Sud-Ouest, Talence, France; LaBRI, UMR 5800, CNRS, Bordeaux INP, Université de Bordeaux, Talence, France; and Institut des Maladies Neurodégénératives, UMR 5293, CNRS, Université de Bordeaux, Bordeaux, France xavier.hinaut@inria.fr

978-1-5090-5069-7/16/$31.00 ©2016 IEEE

ESN, sentences are transformed into a sentence structure (or *grammatical construction*): semantic words (SW), i.e. nouns, verbs and adjectives that have to be assigned a thematic role, are replaced by the *SW* item. The processing of the grammatical construction is sequential (one word at a time) and the final estimation of the thematic roles for each SW are read-out at the end of the sentence.

By processing grammatical constructions and not sentences *per se*, the model is able to bind a virtually unlimited number of sentences to these sentence structures. Based only on a small training corpus, this enables the model to process future sentences with currently unknown semantic words. One major advantage of this neural network language module is that no parsing grammar has to be defined a priori: the system learns only from the examples given in the training corpus. Here are some input/output transformations that the language model performs:

- "Please grasp the cup" → *grasp(cup)*
- "The box is on the right of the cup" → *right(box, cup)*
- "Right of the cup is the box" → *right(box, cup)*

As shown, the system can robustly transform different types of sentences, and even though the last two sentences are quite different (the order of words is different), the system can learn to provide an identical predicate representation. Recently, we have explored the flexibility of the system: we found that it can handle various kinds of representations at the same time. For instance, it allows to use nouns as main elements of a predicate, and use its arguments to fill in adjectives:

- "Find the big red object"
  → *find(object), object(big, red)*

C. Usage of the ROS Module

The proposed model was encapsulated in a ROS module. It is coded in Python language and uses *rospy* library. When running the program, the reservoir model is trained using the given text file and the ROS service is initialized. The training text file contains sentences and corresponding predicates and can be edited easily. Here is one line of a training text file as an example:

- `find box, box blue; find the blue box`

This predicate representation enables one to easily integrate this model into a robotic architecture. If predicates represent multiple motor actions that have to be performed in a particular order, the predicates have to be specified in chronological order:

- `show banana, point box; first show me the banana and then point to the box`
- `show banana, pointing box; before pointing to the box show me the banana`

Once initialized, a request could be sent to the ROS service: it accepts a sentence (text string) as input and returns an array of predicates in real time. With an ESN of 100 units, the training of 200 sentences takes about one second on a laptop computer. Testing a sentence is of the order of 10 ms.

III. Discussion

Previous experiments have shown that the system can be used in a real-world robot scenario (e.g. with the iCub [2] and the Nao [4][5] humanoid robots). In Hinaut et al. [6], it has been shown that the model can learn to process sentences with out-of-vocabulary words. Moreover, we demonstrated that it can learn sentences both in French and English at the same time. To illustrate how the robot interaction works, a video can be seen at `youtu.be/FpYDco3ZgkU` [4][5]. The source code is available at `github.com/neuronalX/EchoRob`.

This ROS module could be employed to process various hypotheses generated by a speech recognition system (like in [5]), then returning the retrieved predicates for each hypothesis, thus, enabling a semantic analyser or world simulator to choose the predicates with the highest likelihood. Preliminary work has shown that the model could be trained fully incrementally [7], we plan to add this feature to the ROS module in further work.

The objectives of this model are to improve HRI and provide models of language acquisition. From the HRI point of view, the aim of using this neural network-based model is (1) to gain adaptability because the system is trained on corpus examples (no need to predefine a parser for each language), (2) to be able to process natural language sentences instead of stereotypical sentences (i.e. "put cup left"), and (3) to be able to generalize to unknown sentence structures (not in the training data set). Moreover, this model is quite flexible when changing the output predicate representations, as we have shown here. From the computational neuroscience and developmental robotics point of view, the aim of this architecture is to model and test hypotheses about child learning processes of language acquisition [8].

Acknowledgment

We thank Iris Wieser for fruitful discussions and ideas, and Fabien Benureau for his very useful feedback.

References

[1] H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148:34, 2001.

[2] X. Hinaut, M. Petit, G. Pointeau, and P. F. Dominey. Exploring the acquisition and production of grammatical constructions through human-robot interaction with echo state networks. *Front. in Neurorob.*, 8, 2014.

[3] X. Hinaut and P.F. Dominey. Real-time parallel processing of grammatical structure in the fronto-striatal system: a recurrent network simulation study using reservoir computing. *PloS one*, 8(2):e52946, 2013.

[4] X. Hinaut, J. Twiefel, M. Borghetti Soares, P. Barros, L. Mici, and S. Wermter. Humanoidly speaking ... In *Video competition, IJCAI, Buenos Aires, Argentina*, 2015.

[5] J. Twiefel, X. Hinaut, M. Borghetti, E. Strahl, and S. Wermter. Using Natural Language Feedback in a Neuro-inspired Integrated Multimodal Robotic Architecture. In *Proc. of RO-MAN*, New York City, USA, 2016.

[6] X. Hinaut, J. Twiefel, M. Petit, P. Dominey, and S. Wermter. A recurrent neural network for multiple language acquisition: Starting with english and french. In *Proc. of CoCo@NIPS workshop*, 2015.

[7] X. Hinaut and S. Wermter. An incremental approach to language acquisition: Thematic role assignment with echo state networks. In *Proc. of ICANN 2014*, pages 33–40, 2014.

[8] M. Tomasello. *Constructing a language: A usage based approach to language acquisition*. Cambridge, MA: Harvard University Press, 2003.