Procedia Computer Science
Volume 71, 2015, Pages 196–201
2015 Annual International Conference on Biologically Inspired Cognitive Architectures

A cognitive neural model of executive functions in natural language processing

Bruno Golosio[1*], Angelo Cangelosi[2], Olesya Gamotina[1], Giovanni Luca Masala[1]

[1] *POLCOMING Department, Section of Engineering and Information Technologies, University of Sassari, Italy*

[2] *Centre for Robotics and Neural Systems, School of Computing and Mathematics, University of Plymouth, United Kingdom.*

[*] corresponding author; email: golosio@uniss.it

Abstract

Although extensive research has been devoted to cognitive models of human language, the role of executive functions in language processing has little been explored. In this work we present a neural-network-based cognitive architecture which models the development of the procedural knowledge that underpin language processing. The large scale organization of the architecture is based on a multi-component working memory model, with a central executive that controls the flow of information among the slave systems through neural gating mechanisms. The system was validated, starting from a *tabula rasa* condition, on a on a corpus of five datasets, each devoted to a thematic group, based on literature on early language assessment, at the level of a preschool child. The results show that the system is capable of learning different word classes, and to use them in expressive language, through an open-ended incremental learning process, expressing a broad range of language processing functionalities.

Keywords: large-scale artificial neural networks, human language understanding, verbal working memory, cognitive architectures, Hebbian learning rule

1 Introduction

Neural-network language models are suitable for modeling the cognitive foundations of language processing and for representing statistical regularities in natural language [1-3]. They have widely been used in natural language processing (NLP), demonstrating superior performances over conventional approaches in next-word prediction and other standard NLP tasks. Recently, deep learning techniques based on recurrent neural networks (RNNs) have been used successfully for several NLP tasks, including speech recognition [4], parsing [5,6], machine translation [7], sentiment

Selection and peer-review under responsibility of the Scientific Programme Committee of BICA 2015

doi:10.1016/j.procs.2015.12.200

analysis of text [8]. Although some of these models are biologically inspired, they are mainly designed as engineering solutions to specific problems in NLP. On the other hand, little work was done to integrate neural models of language into comprehensive cognitive models compatible with current knowledge of how verbal information is stored and processed in the brain. Miikkulainen [3,9] and Fidelman et al. [10] presented a cognitive neural architecture able to parse script-based stories, to store them in episodic memory, to generate paraphrases of the narratives, and to answers questions about them. Their model was tested on a small corpus of nine scripts, each of which consisted of 4-7 sentences. Dominey and Hinaut [11,12] proposed a neural model of brain areas involved in language processing, able to learn grammatical constructions and to generalize the acquired knowledge to novel constructions.

In this work, we present a cognitive neural model aimed at explaining the development of the procedural knowledge that is involved in processing verbal information at the sentence level, combining it with information retrieved from long-term memory, selecting relevant items and planning language production. The purpose of this work is to contribute to understanding the mechanisms that make the human brain able to develop a broad range of language processing skills, starting from a tabula rasa condition.

2 Methods

The model described in this work, called ANNABELL (Artificial Neural Network with Adaptive Behavior Exploited for Language Learning), is a cognitive neural architecture that was designed to help understand the cognitive processes involved in early language development. A detailed description of the model and of the database used for its validation is provided in Ref. [13]. The source code of the software, the User Guide and the datasets used for its validation are available in the ANNABELL web site at https://github.com/golosio/annabell/wiki . The global organization of the architecture is based on a multi-component working memory model [14]. The model comprises four main components, as shown in Fig. 1: a verbal short-term memory (STM), a verbal long-term memory (LTM), a central executive (CE) and a reward structure.

The STM includes a phonological store, a focus of attention, a goal stack and a comparison structure. The phonological store maintains the working phrase, which is either acquired from verbal input or retrieved from LTM. The focus of attention holds up to about four words. The goal stack is a structure for storing goal chunks that contributes to decision-making processes. The comparison structure recognizes similarities among words in the phonological store, in the focus of attention and in the goal stack, and is also used for decision-making processes.

The LTM includes a structure for memorizing the working phrases, and a retrieval structure that uses the focus of attention as a cue for retrieving memorized phrases.

The CE controls all decision-dependent processes. It includes a state-action association system, a set of action neurons and a set of gatekeeper neurons. The state-action association system is a neural network that is trained by a rewarding procedure to associate mental actions to the internal states of the system.

The reward structure is a system that memorizes the sequences of the internal states and of the mental actions performed by the system (state-action sequences) during the exploration phases. When the exploration produces a target output, the reward structure retrieves the state-action sequence and rewards the association between each internal state and the corresponding mental action, by triggering synaptic changes of the state-action association connections.

At the lowest level, the system is entirely composed by interconnected artificial neurons. The learnable connections among neurons are updated by a discrete version of the Hebbian learning rule (DHL rule) [15,16]. The inhibitory competition among groups of neurons is modeled by the *k-winner-take-all* rule [17].

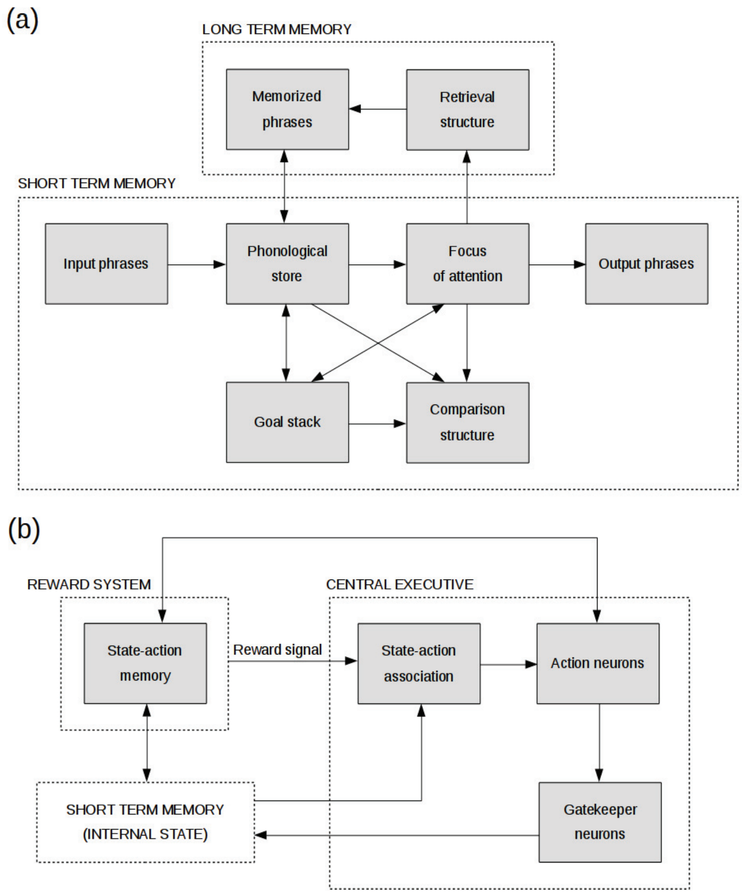

Figure 1. Schematic diagram of the main system architecture. (a) Short-term memory and long-term memory. (b) Central executive and reward structure.

The network architecture is designed in such a way that the system can elaborate phrases using mental actions, which are elementary operations on word groups and phrases that are used, for instance, for acquiring the words of the input phrases, for memorizing phrases, for extracting word groups from the working phrase, for retrieving memorized phrases from word groups through an association mechanism, etc. Such actions are performed by special neurons, called mental action neurons, which can control the flow of signal between different subnetworks. Each mental action neuron can activate one or more gatekeeper neurons, which control the flow of signal between different subnetworks. Neural gating mechanisms play an important role in the cortex and in other regions of the brain [18]. In the ANNABELL model, the gatekeeper neurons are generally fully connected to one or more subnetworks, and they can enhance or suppress the activation state of the postsynaptic neurons, acting in a similar way to a change in the bias signal. In this manner they can control the flow of signal from one part of the system to another.

A key feature of the ANNABELL system that is particularly important for its generalization capabilities is that the learnable connections that are affected by the reward (i.e. the connections of the state-action association system) are connected to action neurons, rather than being directly connected to output words or phrases. In this way, the system learns preferentially to build the output through sequences of elementary operations on word groups or phrases.

The system was implemented on a PC equipped with a high-performance GPU (graphics processing unit) NVIDIA Kepler GK104 having 1536 cores. The current version of the system is composed by 2.1 million neurons, interconnected through 27 million connections.

3 Results and Discussion

The database used for training and validating the system is based on literature on early language assessment. Its sentences have been prepared by treating the system as a child in a virtual social environment. The database is organized in five datasets, which include declarative sentences and interrogative sentences. The first dataset, devoted to the subject people, includes sentences related to the people that belongs to the social environment of the girl impersonated by the system, and it is inspired by the Language Development Survey work of Rescorla [19,20]. The second dataset, called parts of the body, includes sentences related to the definition, location and function of thirty-three body parts. The third dataset, named categorization, uses the animal-world classification for training the system to answer simple questions related to categorization, category hierarchies and combinations of categories and adjectives. The fourth dataset, devoted to the communicative interactions, is based on the Warren-Leubecker corpus [21,22] from the CHILDES database [23]. This corpus contains data transcribed from child-parent conversations. In particular, the session used in this work (the file “david.cha”) is a conversation between a 5-years-and-10-months old child and his mother. The fifth dataset represents a text-based virtual environment where the system can perform simple tasks by means of verbal commands. The system is trained to accomplish the task of bringing objects to a person, which involves the ability to move in the virtual environment from some starting room to the room where an object is located, to execute the command for taking the object, to move back to the starting room and to execute the command for giving the object to the person.

The training procedure is organized in five learning sessions, one for each dataset. Each session is divided in two stages. During the first stage, a set of declarative sentences from the corresponding dataset is presented to the system. As the system does not have any other sensory input, all the information must be provided to it in the form of verbal descriptions. In the subsequent training stage, the teacher trains the system by asking it a set of questions related to the previous sentences. During this stage, the system works in an exploration modality, which combines an association mechanism with partially-random operations on word groups and phrases. The teacher guides the system to retrieve useful associations, and instructs the interface to trigger a reward signal when the system yields a correct answer.

The evaluation of the system performance (test stage) is performed at the end of the learning sessions, after the cumulative training on all five datasets. In this stage, the teacher evaluates the system by asking it a set of questions similar to the ones used during the training stages, and by testing the generalization capabilities of the system, i.e. its ability to elaborate the information provided by the memorized sentences, and to answer questions having a similar structure to those presented during the training stages but involving different nouns, adjectives or verbs. The teacher also validates the linguistic competences of the system in the use of articles, nouns, verbs, adjectives, personal pronouns, possessive pronouns and other word classes. The system output sentences were considered to be valid when they were syntactically and semantically correct and appropriate for the conversation. In the test related to the communicative interaction dataset, the human interlocutor brought out the system in a conversation similar to the one transcribed in the corpus.

The system performance was evaluated through a four-round cross validation. In a single round of the cross validation, the system was trained and evaluated using 1587 input sentences, containing 595 different words, with an average number of 5.6 words per sentence. It produced 521 output sentences, containing 312 different words (expressive vocabulary), with an average number of 4.6 words per sentence.

The percentage of correct output sentences over the total number of requested output sentences, averaged over the four rounds of the cross validation, was 82.4% for the people dataset, 85.3% for the parts of the body dataset, and 95.3% for the categorization dataset. The communicative interaction dataset and the virtual environment dataset are excluded here because they are not suitable for this type of evaluation.

In the session based on the Warren-Leubecker corpus, the total number of tokens (words) used by the system in the output sentences was 111, while those used by the real child in the same conversation were 134. The total numbers of different token types were 75 and 86, respectively. The type/token ratios are close to each other, being 0.68 for our system and 0.64 for the child.

The capacity of the system to generalize acquired knowledge to new situations was tested in two generalization experiments. The first experiment tested the ability of the system to handle learned grammatical constructions with new open-class words. For this purpose, we used an extended corpus of 5352 declarative sentences and 4028 interrogative sentences. This corpus was generated by replacing the open-class words of the three datasets people, parts of the body and categorization, with new, randomly generated words, preserving the correct correspondence among words in the declarative sentences and words in the interrogative sentences. The percentage of correct output sentences over the total number of requested output sentences was 88.1% for the people dataset, 88.3% for the parts of the body dataset, and 76.7% for the categorization dataset. The second experiment evaluated the capacity of the system to generalize acquired knowledge to new grammatical constructions (compositional generalization) on a task of sentence-to-meaning mapping. This experiment was performed on a corpus of 462 distinct grammatical constructions, developed by Hinaut and Dominey [12]. A ten-fold cross validation on this dataset yielded 9.2% meaning error and 36.7% sentence error rate.

4 Conclusion

The validation show that, compared to previous cognitive neural models of language, the model is able to develop a broad range of functionalities, starting from a tabula rasa condition. Those results support the hypothesis that executive functions play a fundamental role for the elaboration of verbal information. Our work emphasizes that the decision processes operated by the central executive are statistical decision processes, which are learned by exploration-reward mechanisms, rather than being based on pre-coded rules.

References

1. Elman JL (1991) Distributed representations, simple recurrent networks, and grammatical structure. *Mach Learn* 7(2-3): 195-225.
2. McClelland JL, Kawamoto AH (1986) Mechanisms of sentence processing: Assigning roles to constituents of sentences. in *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, eds McClelland JL, Rumelhart DE (MIT Press, Cambridge, MA).
3. Miikkulainen R (1993) *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory* (MIT Press, Cambridge, MA).
4. Mikolov T, Zweig G (2012) Context dependent recurrent neural network language model, in *Proceedings of the IEEE Workshop on Spoken Language Technology, SLT 2012*, pp 234-239.

5. Socher R, Lin CC, Ng AY, Manning CD (2011) Parsing natural scenes and natural language with recursive neural networks, in *Proceedings of the 26th International Conference on Machine Learning (ICML 2011)*, vol. 2, pp. 129-136.
6. Socher R, Bauer J, Manning CD (2013) Parsing with compositional vector grammars, in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 455–465.
7. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks, in *Advances in Neural Information Processing Systems 27 (NIPS 2014).*
8. Socher R, Perelygin A, Wu J, Chuang J, Manning CD, Ng AY et al. (2013) Recursive deep models for semantic compositionality over a sentiment treebank, in *Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*.
9. Miikkulainen R (1995) Script-based inference and memory retrieval in subsymbolic story processing. *Applied Intelligence* 5(2) 137-163.
10. Fidelman P, Miikkulainen R, Hoffman R (2005) A Subsymbolic Model of Complex Story Understanding, in *Proceedings of the 27th Annual Meeting of the Cognitive Science Society 2005*.
11. Dominey PF (2013) Recurrent temporal networks and language acquisition—from corticostriatal neurophysiology to reservoir computing. *Front Psychol* 4. 500.
12. Hinaut X, Dominey PF (2013) 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
13. Golosio B, Cangelosi A, Gamotina A, Masala GL (2015) A cognitive neural architecture able to learn and communicate through natural language. *PLOS ONE* (in press, doi: http://dx.doi.org/10.1371/journal.pone.0140866).
14. Baddeley AD (2012) Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology* 63: 1-29.
15. Hebb DO (1949) *The Organization of Behavior* (Wiley & Sons, New York).
16. Haykin S (1998) *Neural Networks: A Comprehensive Foundation* (Prentice Hall PTR Upper Saddle River, NJ, USA, 2nd edition).
17. O'Reilly RC, Munakata Y (2000) *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain* (MIT Press, Cambridge, MA).
18. Boukadoum M, Gisiger T (2011) Mechanisms Gating the Flow of Information in the Cortex: What They Might Look Like and What Their Uses may be, *Frontiers in Computational Neuroscience* 5(1).
19. Rescorla L, Alley A (2001) Validation of the language development survey (LDS): a parent report tool for identifying language delay in toddlers, *J Speech Lang Hear Res* 44(2): 434-445.
20. Rescorla L, Achenbach T (2002) Use of the Language Development Survey (LDS) in a National probability sample of children 18 to 35 months old, *J Speech Lang Hear Res* 45(4): 733-743.
21. Warren-Leubecker A (1982) Sex differences in speech to children. *Unpublished doctoral dissertation* (Georgia Institute of Technology).
22. Warren-Leubecker A, Bohannon JN (1984) Intonation patterns in child-directed speech: Mother-father speech. *Child Dev* 55: 1379-1385.
23. MacWhinney B (2000) *The CHILDES Project: Tools for analyzing talk* (Lawrence Erlbaum Associates, Mahwah, NJ, 3rd edition).