Original scientific paper

Biological and Cognitive Plausibility in Connectionist Networks for Language Modelling

Maja Anđel
Department for German Studies
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
mandel@ffzg.hr

Summary

If we want to explain cognitive processes with means of connectionist networks, these networks have to correspond with cognitive systems and their underlying biological mechanisms in different respects.

The question of biological and cognitive plausibility of connectionist models arises from two different aspects – first, from the aspect of biology – on one hand, one has to have a fair understanding of biological mechanisms and cognitive mechanisms in order to represent them in a model, and on the other hand there is the aspect of modeling – one has to know how to construct a model to represent precisely what we are aiming at. Computer power and modeling techniques have improved dramatically in recent 20 years, so the plausibility problem is being addressed in more adequate ways as well. Connectionist models are often used for representing different aspects of natural language. Their biological plausibility had sometimes been questioned in the past. Today, the field of computational neuroscience offers several acceptable possibilities of modeling higher cognitive functions, and language is among them.

This paper brings a presentation of some existing connectionist networks modeling natural language. The question of their explanatory power and plausibility in terms of biological and cognitive systems they are representing is discussed.

Key words: connectionism, computational neuroscience, language

Introduction

The development of computers in the second half of the twentieth century and new insights in functioning of the human brain, which developed simultaneously, brought many researchers to start thinking of brain as of an extremely sophisticated information processing device. This idea (often labelled as *computer-metaphor*) led further to the emergence of the so-called subsymbolic models of cognitive processes. The first wave of enthusiasm subsided after recognizing limits of perceptrons (Minsky, Papert 1969), but after a while, network architectures were improved (Rumelhart, McClelland, PDP research group 1986) and computers became much faster and more efficient, resulting in an increased interest in artificial neural networks.

From Early Connectionism to Computational Cognitive Neuroscience

Connectionism (term first used in Feldman, Ballard 1982) is

> “[...] an approach to artificial intelligence (AI) that developed out of attempts to understand how the human brain works at the neural level and, in particular, how people learn and remember. (For that reason, this approach is sometimes referred to as neuronlike computing.)”[1]

The term itself refers to the fact that small computing units (neuronlike units) are interconnected through a range of connections whose weights have to be adjusted during the learning process in order to satisfy all the constraints imposed on the network by the learning task.

After 20 years, models as metaphors are closer to their original inspiration, brain, and the term *computational neuroscience*, although not new (cf. Sejnowski, Koch, Churchland 1988), is sometimes preferred, in order to emphasize the increasing similarity:

> “[...] computational neuroscience makes systematic use of mathematical analysis and function of living brains, building on earlier work in both neural modeling and biological control theory.” (Arbib, 2002:11)

From this definition one can see that there are two possible directions for a researcher to choose from: on one side, putting more weight to mathematical analysis, striving for better computational efficiency – which is useful in robotics, or, on the other side, giving more attention to the imitation of biological structures, which leads to building models of cognitive processes. However, recent development has shown that adhering to biologically plausible models does not necessarily mean neglecting computational efficiency, as will be shown in further text.

Biology as inspiration

It is often pointed out that connectionist networks are inspired by biological networks of neurons, and therefore they are also known as artificial neural networks. In the beginnings of connectionism, neurons and populations of neurons were represented as simple computational units, and connections between them as relatively simple mathematical functions. Due to this reductionism, connectionist models have often been criticized in different respects (Fodor, Pylyshin 1988, Pinker, Prince 1988).

[1] Citation from Britannica online, accessed on August 13, 2009.

However, as the time goes on, the mathematical functions and algorithms have become more sophisticated and more adjusted to known biological processes and functions, so the focus of criticism has shifted towards other domains. Recently, their capability to represent cognitive processes without using symbol manipulation has been questioned (cf. Marcus 2003). This debate (symbolic vs. subsymbolic processing) still goes on without an answer (cf. Christiansen, Chater 1999), so there is no intention to give the answer in this paper either, but rather to show what connectionist models are capable of doing today in explaining psycholinguistic processes.

Connectionism and language (history)

Language was one of the cognitive domains covered in the most important volumes dealing with parallel processing in the eighties – the Parallel Distributed Processing by Rumelhart, McClelland and the PDP group (1986). The position of linguistic phenomena in these volumes indicates the importance of language for modelers of cognitive processes. Today, however, the focus has somewhat shifted more towards visual processing and memory. Models of linguistic processing are still present, but their proportion seems to have decreased.

The models presented here are by far not all that deserve to be presented for their importance in development of connectionist networks. They were selected due to the fact that the themes they cover – language morphology, represented by models of English past tense acquisition, and language syntax, represented by models of thematic role assignment – are the themes that appear repeatedly in connectionist models. Nevertheless, all those models bring something new in terms of architecture or algorithms, and therefore they are different in respect to their biological plausibility. Because of their importance for the architecture of linguistic models, descriptions of Elmans models (Elman 1990) are also added.

Models

First models

One of the most debated models in the history of connectionism is the model of English past tense acquisition by Rumelhart and McClelland (1986). With the perceptron learning algorithm (Rosenblatt 1962) they trained the network to connect 430 root verbs with their past tense forms. For the process of learning itself, it was important to imitate the U-shaped curve observed by researchers of child language acquisition by that time (e.g. Brown 1973). With no means to represent the temporal sequence of phonemes, they used a system of so called wickelfeatures, where three subsequent phonemes were encoded as one input block, followed by another block of three phonemes where the phonemes are moved by one and so on. They started with training the network on 10 verbs, and after that they proceeded with all remaining 420 verbs. The procedure yielded the desired learning outcome, with the network being able to connect most of the verbs with their correct past tense form and exhibiting an U-shaped form of the learning curve (showing the process of overregularization), but the procedure itself (training the network first on a very small number of verbs, than suddenly increasing the training set) as well as the system of wickelfeatures were often pointed out as being problematic (Pinker, Prince 1988).

McClelland and Kawamoto (1986) model of assigning thematic roles to words within sentences, i.e. understanding their meaning along with their syntactic structure shows that networks are capable of learning concepts greater than words - sentences. Sentences are presented to the network as strings of words, which are in turn represented as subsets of microfeatures. The architecture is similar to those from Rumelhart and McClelland (1986) – simple two-layered, feed-forward perceptron.

The first models showed that there is a good reason to believe that connectionist models are indeed capable of approximating at least some aspects of human cognitive processes.

From the point of view of biological plausibility, few questions were posed at the time. The first goal – cognitively plausible models – seemed possible to achieve, and the second – biological plausibility – was yet to come.

Some important models of natural language in the past

Elman (1990) proposes a method for encoding temporal sequences, called simple recurrent network (SRN). Instead of encoding sequences in a spatial manner (as seen e.g. in wickelfeatures introduced by Rumelhart and McClelland (1986)), he introduces an additional layer to a feedforward network. The new layer "memorizes" the current step of the system and feeds its contents back into the hidden layer along with the contents of the next step. In this way, the network takes into account what it had learnt in the past - a temporal sequence of elements. Elman validates the method by testing it on four tasks, three of them of linguistic nature - learning a letter sequence, learning to recognize word boundaries, learning to categorize words depending on syntactic features of simple sentences. The network performs successfully on all three tasks. In the first one, it has to learn three short pseudowords - *ba*, *dii* and *guu*. It successfully learns to predict vowels, because they always appear after same consonants. It also learns the length of sequences, since they are fixed for every word. In the second task, the network learns 14 pseudowords of different length. The more elements the network obtains for recognizing the word, the better its prediction for the next element (letter). When the word ends, the networks prediction for the next element is again inaccurate, because it is never sure what the next (randomly picked) word will be. One can say that the network has successfully learned to recognize word boundaries. In the third task, short sentences (two or three words, represented as sparse localist 31-bit vectors) were presented to the network. This time, the network learns to predict possible word(s) to follow and the likelihood of their occurrence. In addition, by analyzing the internal representations for each word in the hidden layer, Elman shows that the network has correctly organized the words into semantic categories, based only on statistical data - their co-occurrence, sequential order and context. However, Elman finds that the categories are not always distinct and clear - some category boundaries seem to be “soft” and implicit.

Trying to improve results obtained earlier by Rumelhart and McClelland (1986) in modeling English past tense, Plunkett and Marchman (1993; 1991) adjust the training procedure in order to make it more similar to the actual input received by children acquiring language. They also start with a smaller set of verbs (initially 20, eventually 500), but the increase in number of verbs is gradual. New verbs are added only after all previous have been acquired. They also try to find out the critical extent of irregularities in the training set that might have an impact on learning and explore the possible differences in sizes of the hidden layer. Their network is a feed-forward perceptron with a hidden layer, using the backpropagation algorithm. Using the new training regime, that was more plausible from the cognitive point of view (more realistic in terms of language acquisition) they obtained better results than Rumelhart and McClelland initially – the U-shaped learning occurred without manipulations of the training set. Many more models on same topic with similar outcomes were made at that time or somewhat later (Daugherty, Seidenberg 1992; Hare, Elman 1992; Hoeffner 1997).

In a recurrent network for sentence processing and decoding, St. John and McClelland (1991) model a system that makes internal representations of entire sentences with their syntactic and semantic properties, similar to cognitive sentence frames made by speakers of natural languages. The network (called the Sentence Gestalt network) could recognize thematic roles for words within sentences, even when they were ambiguous – the network could distinguish thematic roles for words, depending on their variable semantic role within sentences (for possible critique on Sentence Gestalt cf. Plaut, Kello 1999). With four hidden and context layers, the architecture of their model is somewhat more complex than usual for simple recurrent networks. The difference in word representation/encoding is also important, because due to the criticism of Kawamoto and McClelland (1986) model, they avoid to encode words as subsets of microfeatures and use localist representations instead (Waskan 2001).

As much as all the described models represent further improvements for models' similarity with cognitive processes they aim to describe, they all basically rely on the same principle of backpropagation, which was heavily debated for its incompatibility with real biological mechanisms. Furthermore, the question of representations (localist or distributed) was raised – distributed representations, such as microfeatures used by McClelland and Kawamoto (1986), were said to disclose to the network too much information that it should extract (learn) from the data on its own. On the other hand side, the localist representations do not correspond with the biological reality of data processing.

As one can see, in the nineties the models were gradually improving when it comes to their performance in representing cognitive processes. However, their closeness to their biological ideal remained under question mark, with the widely used backpropagation algorithm and localist representations (and some other features).

Today

In the last decade, there were several proposals to improve the problem of biological plausibility. One of them was postulated by O'Reilly (1998). He described six principles that should be followed in order to achieve greater biological plausibility: The first principle, the biological realism, should be the central theme of the cognitive modeling in general. Thereafter, all models should be "constrained and informed by the biological properties" of the brain (O'Reilly 1998:456). It is not enough that models imitate cognitive processes, but they should do so by respecting the biological properties of the brain.

The first principle is followed by three more that describe the ideal network architecture: distributed representations, inhibitory competition and bidirectional activation propagation. It is believed that the cortex uses distributed (more neurons are activated at the same time in order to represent a concept), rather than localist representations (one neuron-one concept), so the networks should do the same. Inhibitory competition between neurons assures that in the process of learning only the most strongly excited activations remain active (and the less excited are inhibited), and therefore enables the network to make fine differentiations between concepts. In other words, it enables the network to successfully distinguish between (even similar) concepts. The bidirectional activation propagation makes it possible to the network to function both bottom-up and top-down, as it is case in human cognitive processes. In the example of reading, we recognize words by recognizing letters (bottom-up), but we can also read a word even if we fail to positively identify one of the letters – if we manage to recognize the entire word, it will help us fill in the gap (top-down). The two remaining principles refer to the learning process – the error-driven task learning and the Hebbian model learning. In the past, the error-driven task learning (supervised learning) was most often implemented as the error backpropagation (Rumelhart, Hinton, Williams 1986). This procedure was criticized, because similar process does not exist in neurobiology. An alternative to it has been proposed as early as in 1987 (GeneRec by Hinton, McClelland 1987), but it was not widely used until 1996 (Leabra, as an improvement of GeneRec, by O'Reilly 1996). It is about settling the network weights in two phases – in the first phase, the network's guess is propagated up to the output layer, and in the second phase the expected outcome (teacher signal) is propagated. Without any backpropagation the difference between the two signals is computed, which represents the network's error, and the weights are adjusted according to this error. The more the network is informed about its errors, the better its guesses become – it learns. Further constraining an error-driven network by Hebbian learning can facilitate the learning and improve the network's results.

In his paper, O'Reilly (1999) discusses all of these principles and the fact, that they are not often combined in models, which makes models less biologically plausible; he acknowledges the fact that some of the principles seem to be in conflict, but also offers ways to overcome the difficulties – not by simplifying, but rather by combining all (or most of) the principles described above.

In addition, he proposes a new algorithm, called Leabra, in which all of his principles were implemented, the biological realism above all:

> "[…] the algorithms they [O'Reilly, Munakata 2000] introduce are constrained by biologically realistic principles, and the resulting models thus incorporate detailed assumptions about such things as membrane potentials, leak currents, or spiking rates." (Cleeremans, Destrebecqz 2003)

O'Reilly and Munakata (2000) describe their models of English past tense and Sentence gestalt, using the Leabra algorithm (as replications of Hoeffner 1997 and St. John and McClelland 1990, respectively). For the English past tense model, they criticize older models for the fact, that it in all of them there is no clear distinction between two levels of analysis that both influence the U-shaped learning – the mechanistic level (mechanic properties of the model itself) and the environmental level (structure of input data). Relying only on the Leabra algorithm, no context layers in the network, their model is mapping from semantics of the words to their phonological shape. Their results are even more consistent with data for human speakers, compared to usual backpropagation models.

The success of replicating the Sentence Gestalt model was comparable to the original, managing to satisfy multiple constraints set by syntax and semantic simultaneously. Some advantages, however, were observed: learning was much faster; Hebbian learning and inhibitory competition, added by the Leabra algorithm made the model more real neuron-like.

Rosa (2004) compared the performance of two connectionist networks trying to solve the same task on the same set of sentences. The task consisted in assigning the thematic roles to words within sentences, fed to the network one at a time. The words were encoded as subsets of distributed microfeatures. The only difference between the networks were their learning algorithms - the first one used the backpropagation algorithm on a simple recurrent network of the Elman type (recurrent connections to the context layer from the hidden layer), whereas the second used the Leabra algorithm, with no need for additional context layers. The training corpus consisted of 364 different sentences as combinations of 30 nouns and 13 verbs, with some verbs allowing for more than one semantic interpretation:

The man hit something.
The stone hit something.

Thematic roles to formal subjects of these two sentences are different – *man* is the agent of the action, and *stone* is its cause. The ability of the network to distinguish between the two shows the network's deeper understanding of syntactic and semantic relations within the sentence.
Comparing the overall performance of the two networks, Rosa concludes that the network using the Leabra algorithm is not only more acceptable from the biological point of view, but also more computationally efficient.

Conclusion

Although scientists are very much intrigued by psycholinguistic and neurolinguistic phenomena, connectionist models, or computational models of cognitive (linguistic) processing are not very numerous. There are many more models trying to explain our processing of visual stimuli or memory, but the language models are not very common, despite their position in the early modeling literature. Most of the existing models deal with phonetic and/or phonological phenomena, widely used for robotic purposes (voice recognition, simple text/commands recognition). Only few tackle the language in its complexity (such as in Rohde 2002)
Today, the field of the computational neuroscience offers acceptable ways of modeling and exploring higher-level cognitive processes, and can therefore give us valuable insights in the core of many (psycho)linguistic processes.
The art of connectionist networks, their architecture and algorithms have evolved since 1986, so that many of their shortcomings pointed out by watchful observers are no longer valid (such as in Fodor, Pylyshin 1988; Pinker, Prince 1988). All models described in the previous chapter make use of a new algorithm that is more consistent with biological properties of neurons and populations of neurons. Given the fact that Leabra is not the only algorithm of this kind (see O'Reilly 1998), but was only chosen to demonstrate the facts, one can say that connectionist models today are indeed much closer to the biological systems that they were inspired by, than they were only two decades ago.

References

Arbib, Michael A.. The handbook of brain theory and neural networks. Cambridge, MA: MIT Press, 2002

Brown, Roger. A first language. Cambridge, MA: Harvard University Press, 1973

Christiansen, Morten; Chater, Nick. Connectionist natural language processing. The state of the art. // *Cognitive Science.* 23 (1999), 4; 417-437

Cleeremans, Axel; Destrebecqz, Arnaud. Harder, Better, Stronger, Faster: A review of Computational Explorations in Cognitive Neuroscience. // *European Journal of Cognitive Psychology*. 15 (2003), 3; 474-477

Daugherty, Kim; Seidenberg, Mark S.. Rules or connections? The past tense revisited. // *The Proceedings of the 14th Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1992, 259-264

Elman, Jeffrey L. Finding structure in time. // *Cognitive Science.* 14 (1990); 179-211

Feldman, Jerome A.; Ballard, Dana H.. Connectionist models and their properties. // *Cognitive Science.* 6 (1982), 3; 205-254

Fodor, Jerry A.; Pylyshin, Zenon W.. Connectionism and Cognitive Architecture: A Critical Analysis. // *Cognition.* 28 (1988); 3–71

Hare, Mary; Elman, Jeffrey L.. A connectionist account of English inflectional morphology: Evidence from language change. // *The Proceedings of the 14th Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1992, 265-270

Hinton, Geoffrey E.; McClelland, James L.. Learning Representations by Recirculation. // *Neural Information Processing Systems /* Anderson, D. Z. (ed.). New York: American Institute of Physics, 1987, 358-366

Hoeffner, James H.. Are rules a thing of the past? A single mechanism account of English past tense acquisition and processing. Unpublished doctoral dissertation, 1997

Marcus, Gary F.. The Algebraic Mind: Integrating Connectionism and Cognitive Science. Cambridge, MA: MIT Press, 2003

McClelland, James. L.; Kawamoto, Alan H.. Mechanisms of sentence processing: Assigning roles to constituents of sentences. Chapter 19. // *Parallel distributed processing: Explorations in the microstructure of cognition. Volume II.* / McClelland, J. L.; Rumelhart, D. E.; the PDP research group (ed.). Cambridge, MA: MIT Press, 1986, 272-325

Minski, Marvin L.; Papert, Seymour A.. Perceptrons: An Introduction to Computational Geometry. Cambridge, MA.: MIT Press, 1969

O'Reilly, Randall C.. Biologically Plausible Error-driven Learning using Local Activation Differences: The Generalized Recirculation Algorithm. // *Neural Computation.* 8 (1996), 5; 895-938

O'Reilly, Randall C.. Six principles for biologically based computational models of cortical cognition. // *Trends in Cognitive Sciences*. 2 (1998), 11; 455-462

O'Reilly, Randall C.; Munakata, Yuko. Computational explorations in cognitive science. Cambridge, MA: MIT Press, 2000

Pinker, Steven; Prince, Alan. On Language and Connectionism: Analysis of a parallel distributed processing model of language acquisition. // *Cognition.* 28 (1988); 73-193

Plaut, David C.; Kello, Christopher T.. The Emergence of Phonology From the Interplay of Speech Comprehension and Production: A Distributed Connectionist Approach. // *Emergence of Language* / MacWhinney, B. (ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, 1999, 381-415

Plunkett, Kim; Marchman, Virginia. U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. // *Cognition.* 38 (1991); 43–102

Plunkett, Kim; Marchman, Virginia. From rote learning to system building. // *Cognition.*, 48 (1993); 21–69

Rohde, Douglas L. T.. A Connectionist Model of Sentence Comprehension and Production. PhD thesis, School of Computer Science. Pittsburgh, PA: Carnegie Mellon University, 2002

Rosa, Jose L. G.. A Biologically Motivated and Computationally Efficient Natural Language Processor. // *Lecture Notes in Computer Science 2972: Advances in Artificial Intelligence: Proceedings of the Third Mexican International Conference on Artificial Intelligence.* / Monroy, R. et al. (eds.). Heidelberg: Springer Verlag, 2004, 390-399

Rosenblatt, Frank. Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. New York: Spartan, 1962

Rumelhart, David E.; Hinton, Geoffrey L.; Williams, R.. Learning internal representations by error propagation. // *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I.* / McClelland, J. L.; Rumelhart, D. E.; the PDP research group (ed.). Cambridge, MA: MIT Press, 1986, 318-368

Rumelhart, David E.; McClelland, James L.; PDP research group. Parallel distributed processing: Explorations in the microstructure of cognition. Cambridge, MA: MIT Press, 1986

Rumelhart, David E.; McClelland, James L.. On learning the past tenses of English verbs. Chapter 18. // *Parallel distributed processing: Explorations in the microstructure of cognition. Volume II.* / McClelland, J. L.; Rumelhart, D. E.; the PDP research group (ed.). Cambridge, MA: MIT Press, 1986, 216-271

Sejnowski, Terrence J.; Koch, Christoph; Churchland, Patricia S.. Computational Neuroscience. // *Science, New Series*. 241 (1988). 4871; 1299-1306

St. John, Mark F.; McClelland, James L.. Learning and applying contextual constraints in sentence comprehension. // *Artificial intelligence*. 46 (1990); 217-257

Waskan, Jonathan A.. A critique of connectionist semantics. // *Connection Science*. 13 (2001). 3; 277-292