DIVISION OF COMPUTER SCIENCE

Elman's Model of Natural Language Acquisition

Caroline Lyon

Technical Report No.158

April 1993

Elman's model of natural language acquisition

Caroline Lyon

April 1993

1 Introduction

Elman first aims to give an account of natural language showing how a connectionist system can model its significant features. Using the same network and he then goes on to model the acquisition of language by children. This work is described in two papers [1, 2]. His research has been influential outside his specialised field; for example, it figures in a New Scientist feature suggesting that connectionist models can represent childrens' language learning. [3].

Elman's language domain is a very restricted subset of natural language, consisting of 23 words. In his first paper Elman speaks of "studying artificial languages, such as the present one" [1, p.220]. Though he is processing words with a semantic interpretation, these elements are used as symbols in a formal language. Elman's work has already been taken up by others, who have taken forward his ideas on recurrent nets [4, 5, 6, 7] but their objectives involve modelling regular or context free grammars, rather than natural language grammar.

I suggest that Elman makes a significant contribution to connectionist modelling of artificial languages, but that it is premature to extend this to natural language. If this objection is upheld, then it would undermine his theory that his model is "plausibly relevant to [the] learning [capability] characteristic of children."[2]

Since natural language is so hard to model it seems that we have to start with a limited sub language, or (in the inclusive sense) aim to model only certain selected features. However, this should be quite explicit. It is necessary to explore what are the key features of natural language, so that the limitations of the model can be seen, and its objectives made clear. We need to investigate whether our initial simplifications are justifiable and might later be adapted to a more realistic model.

In his second paper [2] Elman focusses on the parallels between effective training methods for neural networks and language learning in children. He shows how incremental training is effective (an approach also described elsewhere [8, 9].) The language that is learnt, however, is the artificial language useded in his first paper.

2 Empirical significance of Elman's model

2.1 Overview

In his earlier work Elman made a significant contribution to modelling sequential input with a connectionist system [10]. He employs the same technique here, using a recurrent network. By saving a state at step n, and then feeding it back at step $n+1$ together with the new input some of the context is captured.

The language domain is a set of 23 words, counting derivations such as "cat" and "cats" separately. There are 10 nouns, 12 verbs and one relative pronoun. These words are combined according to the rules of a simple phrase structure grammar, with 7 productions and 2 extra restrictions. The first restriction specifies noun and verb agreement. The second lists those verbs that must have, may have and cannot have a direct object.

Elman's grammar is a hierarchical structure, as is natural language, and can model some recursive features. He also addresses the issue of context sensitivity in natural language. There are dependencies between adjacent words, between distant words in the same clause, and between words in different clauses.

Elman's model can represent :

- Entry and exit from within-clause state
- Embedded clauses (not more than 2 levels deep).
- Distant agreement between main subject and verb
- Detection of subject or object role of relative pronoun
- Consequent agreement of verb inside clause either with subject of clause or with subject of main sentence

Thus these sentences are grammatically correct within his domain:

Boys who girls who dogs chase see hear.

Dog who chases cat sees girl.

Dog who boys feed sees girl.

While

*Dog who boys feeds see girl

is incorrect.

However, though Elman models these features of natural language, there are a number of other shortcomings.

2.2 The prediction based method

Elman uses a prediction task as a basis for investigating the structure of language. He trains a neural net on a corpus of strings and then tests this by predicting word and word category sequences in unseen strings. Though this is an appropriate way of investigating the structure of a small artificial language it only plays a supporting role in the natural language field. One of the characteristics of natural language is its open ended nature: the non-deterministic element has to be modelled. This is much noted in work on natural language - from theoretical linguistic analysis [11] to the practical development of habitable speech processors coping with 150 ways of saying "yes" or "no" [12]. Indeed, Elman himself starts his first paper [1, p. 195] by identifying three principal issues in natural language processing , one of which is the problem of accommodating the "open ended nature of language".

Prediction as a tool for investigating natural language structure is usually better taken as a one-to-many mapping from a given state onto possible successor states. It would be more appropriate to assess the grammaticality of the strings : at any point in a sentence there may be more than one grammatically correct word or word category that can follow.

Another source of information that can be tapped is the occurrence of collocations - words that are commonly found together. For instance "last year" is a more likely combination than "lost year". However, predictions based on this sort of information cannot be developed in a 23 word domain.

2.3 Compositionality

Natural written language can be represented as strings of words, whose order is significant, partitioned into sentences. Sentences have structure, and the elements out of which this structure is composed are phrases, that is sequences of one or more words that have a certain function. A necessary first step in natural language modelling is to represent these phrases, the initial building blocks or constituents, which are strings of variable length.

Among the foremost critics of connectionist modelling of natural language are Fodor and Pylyshyn [13]. They focus on this issue of compositionality, that is, chunking items together into meaningful groups, as one of the major obstacles to effective connectionist processing.

Elman's system is based on a very simple phrase structure grammar augmented by a context sensitive restriction, which does not currently model variable length constituents. The constituents he produces are of certain fixed lengths. There can be several layers of embedded clauses, but each layer is only one word long. Whether this model could be extended to accommodate compositionality without spoiling the function of the neural net needs further investigation.

2.4 Part of speech ambiguities

A phrase structure of grammar is expressed by allocating words to part-of-speech classes or tags.There are many different tag set systems, ranging in size from those with 2 classes to those with over 200. Many classes have fuzzy boundaries. Many words can belong to more than one class, and this ambiguity seems to be an intrinsic feature of all natural language outside some very limited sub languages [14]. Elman has a 4 class system, but does not currently model this ambiguity. In his choice of 12 verbs, 9 also have another part-of-speech tag within his system, 10 would have if a class of adjectives were included too. Each word is given a single part-of-speech tag - for instance the word "chase" is declared a verb, not noun or verb.

2.5 Generalisation

If this work is to have empirical significance it should be carried out in a wider domain than 23 words that have been ascribed unambiguous tags.

The training set for the net consisted of 50,000 sentences generated by the grammar, apparently all different [2]. Given the restrictions of the language domain it is reasonable to suppose that every word occurred in every possible syntactically correct position, so the test data would not be generalising to new sentences [15].

3 Further developments

3.1 Language processing and state machines

Elman takes issue with the "building block" metaphores that are often used to describe the structure of language. He points out that in his scenario "words are not building blocks as much as they are cues which guide the network through different grammatical states". These two approaches are not inconsistent, and Elman's analogy is illuminating. Building blocks could suggest homogenous elements, whereas the constituents of natural language are of variable length. If a network moves into a within-clause state, for instance, this is analogous to a state machine entering a stack of indeterminate length.

One way of modelling this with a connectionist system is to use a recurrent neural network, very similar to Elman's, with an external stack [16, 6]. Two of the output nodes of the net represent push and pop functions, so that certain states can trigger the operation of the stack. Its length is not determined in advance. These recurrent Neural Network Pushdown Automata (NNPDA) can model simple context free Type 2 grammars.

With its current restrictions Elman's model is derivable from a Type 3 grammar that can be recognised by a non-deterministic finite state automaton. All the possible combinations of allowable sequences could be listed, and a path taken through the lattice. A larger, more realistic model would typically still have a finite vocabulary and finite sentence length, and it might be conceptually possible to continue with such a model. However, since such grammars quickly become computationally intractable, it would be useful to try to develop a model based on a grammar higher in the Chomsky hierarchy [17].

4 Conclusion

Elman has made a significant contribution to the development of connectionist processing techniques. His work with recurrent networks is a particularly important advance. However, the claim that his work is relevant to a model of the acquisition of language by children should be treated with caution. The language he processes differs from natural language in a number of essential ways, and whether it can be extended to a more realistic model remains to be seen. Meanwhile, his work can be seen as an important contribution to processing artificial language.

References

[1] J L Elman. Distributed representations,simple recurrent networks and grammatical structure. *Machine Learning*, 1991.

[2] J L Elman. Incremental learning, or the importance of starting small. Technical report, Center for research in language, University of California at San Diego, CA 92093-0126, 1991.

[3] D Shanks. Breaking chomsky's rules. *New Scientist*, 1993. 30 January issue.

[4] C L Giles and C B Miller. Extracting and learning an unknown grammar with recurrent neural networks. In S J Hanson J E Moody and R P Lippmann, editors, *Advances in neural information processing systems.* Morgan Kaufmann, 1992.

[5] C L Giles et al. Learning and extracting finite state automata with second order recurrent neural networks. *Neural Computation*, 1992.

[6] C L Giles S Das and G Z Sun. Learning context free grammars. In *Proceedings of the Cognitive Science Society*, 1992.

[7] S T Chakradhar M W Goudreau, C L Giles and D Chen. First-order vs second-order single layer recurrent neural networks. Technical report, Princeton University and NEC Research institute, Inc, 1992.

[8] P J Wyard and C Nightingale. A single layer higher order neural net and its application to context free grammar recognition. *Connection Science*, 4, 1990.

[9] B Widrow and Lehr. 30 years of adaptive neural networks. *Proceedings of IEEE*, September 1992.

[10] J L Elman. Finding structure in time. *Cognitive Science*, 1990.

[11] N Chomsky. *Rules and Representations*. Blackwell, 1980.

[12] A L Gorin et al. On adaptive acquisition of language. *ICASSP*, 1990.

[13] J A Fodor and Z W Pylyshyn. Connectionism and cognitive architecture: a critical analysis. *Cognition*, 1988.

[14] G E Barton, R C Berwick, and E S Ristad. *Computational Complexity and Natural Language*. MIT Press, 1987.

[15] R F Hadley. Compositionality and systematicity in connectionist language learning. Technical report, Simon Fraser University, Center for Systems Science, 1993.

[16] C L Giles et al. Higher order recurrent networks and grammatical inference. In D S Touretzky, editor, *Advances in neural information processing systems*. Morgan Kaufmann, 1990.

[17] S M Lucas. New directions in grammatical inference. In *IEE Colloquium on Grammatical Inference*, 1993.