1994 International Symposium on Speech, Image Processing and Neural Networks, 13-16 April 1994, Hong Kong

A Neural Network model for Acquisition of Semantic Structures

*Samuel W.K.Chan   James Franklin**

School of Computer Science and Engineering
*School of Mathematics
University of New South Wales
New South Wales   Australia 2033

Abstract

Recent research suggests that natural language processing (NLP) can be profitably viewed in terms of the spread of activation through a neural network. However, since the critique by Fodor (1988) of the style of connectionist representations, one of the biggest challenges facing proponents of connectionist models of NLP is the rich structures of language. As models of NLP, neural network systems must exhibit the properties of compositionality and structure sensitivity. This paper describes a neural network model in which simple recurrent network and recursive auto-association memory are combined to acquire the semantic structures from sentence constituents. This imposes no prior limit on sentence structures. The model can be viewed as a tool of conceptual acquisition and generalization extraction in language understanding.

1 Introduction

Since the early 1980's, there has been increasing interest in and enthusiasm for neural network researches. This excitement is partly due to indications that it may be possible to build computers with hardware that resembles the human brain. It is also in part due to a common language that neural networks offer to researchers from signal processing, artificial intelligence, neuroscience, cybernetics and psychology as well as linguistics. The commonness of the language is not to be underestimated as a major tool in all disciplines of cognitive science researches. It is the reason why there has been increasing interest in applying connectionist approaches in language analysis recently (Cottrell 1988, St. John and McClelland 1990, Jain 1991, Miikkulainen et al 1991). However, there are significant problems to be surmounted when treating natural language processing in a connectionist framework. One of the aspects of language processing that has been problematic for the connectionist community is that natural language is recursive and it requires complex, structure-sensitive operations that have traditionally been modelled in the symbolic paradigm. This paper presents a neural network model for acquisition of semantic structures from sentence constituents. The goal of this research has been to develop a model that converts a simple sentence into a conceptual representation of the event that the sentence describes. Specifically, it is a model that converts the constituent phrases of a sentence into a representation of a group of conceptual categories.

2 Background

Jackendoff (1972) proposes that there is a set of basic conceptual or ontological categories: Thing, Event, State, Action, Place, Path, Property and Amount (see also Keil, 1979). Jackendoff argues that conceptual abstraction of semantic information does not develop arbitrarily but along a given, predictable path; a developmental path that starts with tangible perceptual predicates (e.g. spatial, causative) which fall under these categories. There is also a set of conceptual formation rules that combine them into more complex concepts. For example, an event may consist of a thing moving along a path. The function relating them is called GO. The first argument of GO, the moving EVENT, is traditionally called the theme. The conceptual formation rule stating that an event can consist of an entity moving along a path is shown in follows.

$$\text{EVENT} \rightarrow [_{\text{event}}\text{GO(THING,PATH)}]$$

In a tree-structure notation, the sentence

(*)   "John goes into his office"

is represented semantically as in figure 1. This representation is based on the idea that semantic structure for a sentence as a whole is built out of the semantic structures of the individual words composing it. This structure forms a hierarchy by naming, in each node, a reference object from the eight conceptual categories and a function, such as TO or IN, in figure 1. This semantic structure representation is crucial for language learners, especially the first language learners, as Pinker (1984) describes that one of the strategies available to language learners involves a sort of template matching of argument to syntactic position.

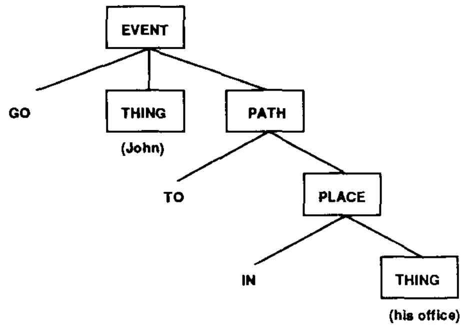

Figure 1: Semantic Structure of Sentence (*)

There are canonical configurations which are default mappings and non-canonical mappings for the exception. Parsing a sentence is often thought of as predicting each successive constituent from those already analyzed on the basis of syntactic rule. However, there are many possible combinations of semantic roles in different sentences. It is unrealistic to enumerate all the possibilities in sentence understanding. In our work, a neural architecture, which can serve to construct and manipulate semantic structures from syntactic constituents, is demonstrated.

3 Description of the Model

A neural network, called the Network for Acquisition of Semantic Knowledge (NASK), combines recurrent network (SRN) and recursive auto-association memory (RAAM) to produce a parser with no prior limit on the length of sentences or the depth of the resulting structures. The NASK architecture consists of two interconnected modules, a parsing subnetwork which is dedicated to parsing whilst a categorization subnetwork is for syntactic-to-semantic transformation. Like all other neural network models, these two consist of networks of interconnected processing units. The detail architecture of NASK is discussed in following sections.

3.1 Encoding Structures

One of the aspects of natural language processing that has been problematic for the connectionist community is that natural language is recursive and it requires complex, structure-sensitive operations that have traditionally been modelled in the symbolic paradigm. Pollack (1990) has developed a subsymbolic model, called Recursive Auto-Association Memory (RAAM), that provides a direct counterexample to the claims that subsymbolic models cannot exhibit useful compositional structure. The basic purpose of RAAM is to allow familiar recursive data structure such as trees and lists to be encoded into distributed representations suitable for processing by connectionist networks. The general architecture of a RAAM is a three-layer feedforward network of processing units and the standard backpropagation algorithm (Rumelhart et al 1986) is employed for learning. The appropriate sets of terminal representations are given as inputs, and the network is trained to reproduce the same representation in the output layer. The hidden layer is thus forced to develop a compressed representation of original inputs. Thus, it produces fixed-length distributed patterns of continuous numerical values in the hidden layer which encode compositional structure implicitly. RAAM can encode general tree structures of variable depth and fixed branching size into fixed-length distributed representations.

Figure 2 shows a RAAM architecture which is used to encode structures in our design. It has 3 branches, each of n units, which are compressed into n units in the hidden layer.

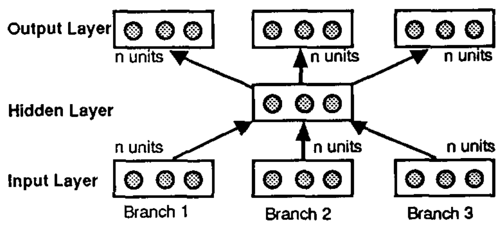

Figure 2: A RAAM Architecture

Representations of the semantic structures of sentences are needed before training with the RAAM in figure 2. They are represented in a fixed branching size 3 as shown in figure 3. Note that dummy object ('NIL') is used to fill in the spaces where only two of the three branches descending from a node are required.

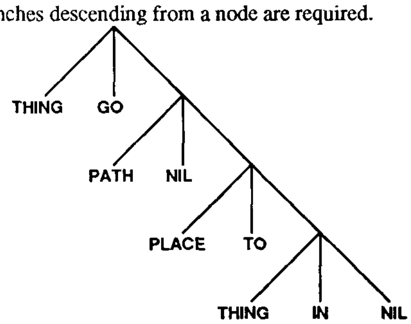

Figure 3: Representation of the Semantic Structure of Sentence (*) in RAAM

3.2 Parsing Subnetwork

In text understanding, the input is structured in time, and thus the behaviour of a system cannot be determined solely on the basis of the current input element. Some sort of memory is required for previous elements in a sentence to be combined with the current element. A simple recurrent network (SRN) introduced by Elman (1991) has been shown to have an ability in predicting successive elements of a sentence. In a recurrent network, the activation levels of the hidden units at time t-1 are presented as input to the network at time t, in addition to the conventional input. This gives the network a history or context, so that the input at time t can be processed in light of what inputs have been presented at earlier times.

In our parsing subnetwork, the parsing technique is based on a combination of SRN and RAAM. The task of the SRN is the construction of a RAAM representation of the syntactic structure of the input sentence. In order to train the SRN, a necessary first step is to create the relevant RAAM representations, as target patterns, of the syntactic structures of the input sentences as described in section 3.1. When parsing, the parsing subnetwork is started with all the context units which are set to a predetermined null value (e.g. 0.5). The encoding of words, which is stored in the lexicon as shown in figure 4, is presented to the subnetwork and activation is fed forward through the network. For each subsequent word in the sentence, the lexical encoding for the word is presented as input, along with the hidden unit activation values from the previous step. After the last word of the sentence has been presented as input, the activation of the output layer units has encoded the representation of the entire sentence.

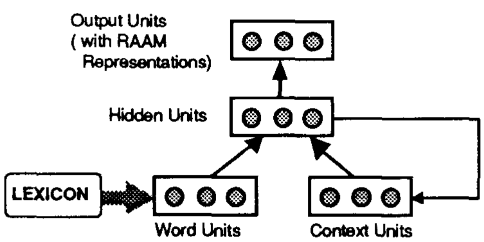

Figure 4: NASK: Parsing Subnetwork

3.3 Categorization Subnetwork

The syntactic-to-semantic transformation takes place in the categorization subnetwork, by operating directly on the compressed distributed representations of sentences. The categorization subnetwork is another simple three-layer feedforward network and makes use of supervised learning as shown in figure 5. The network has no knowledge pre-wired into it which showed for each semantic role differs. The output layer consists of 20 units, the size of the compressed representation developed by the RAAM. The categorization subnetwork takes the output from the parsing network as well as the verb of the input sentence from the lexicon. During training, the distributed representation of the semantic structure of the input sentence is used as its target pattern. Using the RAAM as trained above, the categorization subnetwork is trained to transform directly from the syntactic distributed representation to the semantic one.

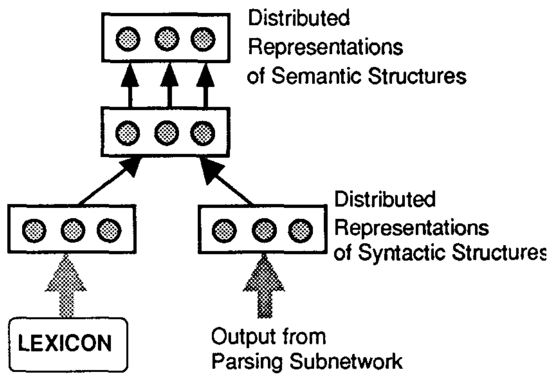

Figure 5: NASK: Categorization Subnetwork

4 Training Method and Results

Determining an appropriate number of encoding units, the value of n in figure 2, is not as straightforward. The number should be large enough to allow the RAAM ample space to successfully compress and reconstruct trees of the required depth, but small enough to allow useful generalizations to develop. In this experiment, the number of the encoding units is chosen by a trial-and-error process, the hidden layer size is set to 20 units. In addition, to represent words as input to a RAAM, each word is randomly assigned with a code. A simple localist representation is used. All of them have no microsemantics and they are stored in a lexicon. The representations of all words are 20 units in all, with 8 units are reserved and never used for the input representations. These 8 units are used to provide the extra spaces which are used by the distributed representations formed by RAAM.

In this small scale experiment, only 100 sentences, with their corresponding semantic structure representations, from the training corpus are encoded using RAAM. There is no restriction on the length or depth of the sentences. To train the parsing subnetwork, the training sentences are presented in a random order. The network is trained for 5,000 epochs. Investigation has shown that the networks converge best if regimes of decreasing learning rate are used. In this experiment, the initial learning rate is set to 0.2 and gradually decreases to 0.02 for every 1,000 epochs. At the end of training, only 4% of output units have values deviated from their target values by 0.05. NASK takes 1000-1500 epochs to reach a rough asymptote for error values. NASK trains to encode the semantic structures for all sentences correctly except the most atypical or extremely deep sentences.

The same training corpus is used in the categorization subnetwork. The distributed representations of syntactic-semantic pairs are arranged as input-output pairs in the categorization subnetwork. On every training cycle, the encoding RAAM representation of syntactic structure of a sentence is input into the categorization subnetwork whilst the desired output of the network is the encoded RAAM representation of the corresponding semantic structure. In the network, the learning rate and momentum are set to 0.1 and 0.9 respectively and is trained for 2,000 epochs.

As a test of generalization, 60 sentences are selected and parsed in the parsing subnetwork. The output is fed into the categorization subnetwork yielding a new distributed representation for each sentence. The representations can be decoded back to their semantic structures. As a result, 34 out of 60 can be decoded back to their corresponding semantic structures. This generalization rate is satisfactory.

5 Conclusions

Connectionism seems to supply the missing ingredient in interactive models of language comprehension. The neural network approach offers advantages over classical parsers in terms of noise tolerance. It has been shown to exhibit functional compositionality and allow structurally sensitive operations. These properties can be applied in acquisition of semantic structures in natural language understanding, as shown in the paper.

References

Cottrell, G.W. (1988). A model of lexical access of ambiguous words. In: Small, S.I. and Cottrell, G.W. (Eds.) *Lexical Ambiguity Resolution.* Morgan Kaufmann Publishers, 179-194.

Elman, J.L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, **7**, 195-225.

Fodor, J.A. and Pylyshyn, Z. (1988).Connectionism and cognitive architecture: A critical analysis. *Cognition*, **28**, 3-71.

Jackendoff, R.S. (1972). *Semantic Interpretation in Generative Grammar.* Cambridge, MA, MIT Press.

Jain, A.N. (1991). Parsing complex sentences with structured connectionist networks. *Neural Computation*, **3**, 110-120.

Keil, F.C. (1979). *Semantic and Conceptual Development: An Ontological Perspective.* Cambridge, MA, Harvard University Press.

Miikkulainen, R. and Dyer, M.G. (1991). Natural language processing with modular PDP networks and distributed lexicon. *Cognitive Science*, **15**, 343-399.

Pinker, S. (1984). *Language Learnability and Language Development.* Cambridge, MA, Harvard University Press.

Pollack, J.B. (1990). Recursive distributed representations. *Artificial Intelligence*, **46**, 77-105.

Rumelhart, D.E. and McClelland, J.L (1986). *Parallel Distributed Processing*, Vol. 1 & 2. Cambridge, MA, MIT Press.

St. John, M.F. and McClelland, J.L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, **46**, 217-257.