Learning Syntactic Frames with Simple Recurrent Networks

Rutvik Desai (RUDESAI@Indiana.Edu)
Computer and Cognitive Science; Indiana University
Bloomington, IN 47505 USA

Introduction

Syntactic bootstrapping and semantic bootstrapping are often seen as rival hypotheses for learning the meaning of verbs. While both theories have different starting points, both credit the child with sensitivity to syntactic information in the input. Hirsh-Pasek and Golinkoff (1997) conducted experiments to examine whether young children are capable of exploiting the syntactic frames in which the verbs are embedded, by testing whether they can distinguish between transitive and intransitive sentence frames.

Here, I present a preliminary connectionist model which uses a Simple Recurrent Network (SRN) (Elman, 1990) to learn to classify similar frames.

A brief, simplified description of the experiments follows.

The Experiments

Hirsh-Pasek and Golinkoff (1997) used an experimental setup where two television screens, set several feet apart, displayed two different visual stimuli. The child sat in the mother's lap, in front of the two screens. Recorded linguistic stimulus was played concurrently with the visual stimuli from a speaker in the middle of the two screens. Two characters, Big Bird (BB) and Cookie Monster (CM), were used throughout the trials. Half of the subjects heard transitive sentences of the form "Look, BB is turning CM!", while the other half heard intransitive sentences such as "See, BB and CM are turning!". Causal and non-causal events corresponding to the verb were shown on the two screens to both the groups. Familiar verbs, such as "turn" and "bend," as well as unfamiliar ones such as "glorp" and "blick" were used in different sentences. The mean visual fixation times were measured for each group by a hidden video camera which recorded the child's responses. If the children comprehend the sentence, they are expected to look at the matching screen for a longer period of time on average.

Twenty-four-month-old subjects looked at the matching screen for a significantly more time than the non-matching screen, implying that they were able to use the syntactic cues in the linguistic input to distinguish between transitive and intransitive frames.

The Network

A SRN was constructed to distinguish between the two syntactic frames, "John is turn ing David." and "John and David are turn ing." Five nouns (which I call John, Mike, David, Linda, Shelly) and five verbs (turn, bend, tickle, flex, twist) were used for generating the training dataset. Using these in the three underlined positions in the frame, a total of 125 sentences of each type were generated. Then end of each sentence was marked with a separate word, STOP. Each word was encoded as an 18-bit vector, with one bit turned on.

If the task of the network is just to distinguish between these two sentences, it can be done in a trivial way, *e.g.*, by checking for the presence of a particular word, or by considering the length of the sentence. To force the network to attend to the structure of the sentence, it was also required to identify the two nouns and the verb in the sentence. In addition, two other sentences of the form "Where are John Mike and David." and "Is John together with David." were added to the training set, to be classified as "other." The choice of these sentences is largely arbitrary, and they are simply supposed to provide some representation to all the other sentences that the children hear with words such as "is" and "are."

The SRN had 18 input nodes, 5+5+5+2=17 output nodes (five nodes for identification of each of the two nouns and the verb, and two for classification), 20 hidden nodes, and 20 context nodes. A learning rate of 0.05 and no momentum were used. The context units were reset after each sentence.

Results and Discussion

The network was able to achieve a near-perfect classification and identification accuracy on the training set after 500 epochs, in each of the several runs with different initial weights. The known verbs were replaced with unknown verbs (generated by turning on multiple bits in the input vector). In all cases, the network correctly identified the sentence type and the nouns. Thus, the network was able to use syntactic information in the frames to distinguish between transitive and intransitive sentences.

The experiments like those described above show that children are sensitive to syntactic cues, and they *can* use this information to infer the meaning of unknown verbs. The question remains whether they actually *do*. The network described here is able to produce the correct responses in spite of ignoring semantics completely. Children could also be relying on syntactic cues to distinguish between sentences, without actually deriving any meaning for the unknown verb.

References

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Hirsh-Pasek, K. and Golinkoff, R. M. (1997). *The origins of grammar: Evidence from early language comprehension.* Cambridge, Mass.: MIT Press.