Context Effects and Learning of Hierarchical Compositional Structure

Karl Pfleger (KPFLEGER@CS.STANFORD.EDU)
Computer Science Department, Stanford University
Stanford, CA, USA

Introduction

Hierarchical compositional structure, in which high level entities represent aggregations of lower level entities, is ubiquitous in people's daily lives (eg, music, spatial configurations, sequences of events). Natural languages, for instance, consist of letters, which together form words, which constitute phrases, etc. Compositional hierarchies (CHs) also play an important role in many successful AI systems.

McClelland and Rumelhart (1981) introduced the landmark Interactive Activation (IA) model, a neural network model of context effects in perception that encoded an explicit CH into a symmetric, recurrent (relaxation-style) neural net, or constraint-satisfaction neural net (CSNN). The network had nodes representing single letters and 4-letter words (I omit the feature nodes). Hand-crafted by the authors, the model did not learn. Despite this model's significant success as a cognitive model, and despite subsequent work on learning algorithms for related CSNNs (eg, Boltzmann Machines), mechanisms for learning this type of network structure have remain unexplored.

I am examining methods for discovering compositional structure bottom-up through repeated composition, or chunking, of lower level entities. I concentrate exclusively on mechanisms that make use only of observed raw primitives, rather than any domain theory or task specific information. Purely data-driven aggregation can be performed by chunking frequently occurring combinations of lower level entities.

This structure leads naturally to smooth integration of bottom-up and top-down processing, and can be used to predict future observations, filter noise, fill in missing or ambiguous entries from context (as in the IA model), or detect anomalies or errors and suggest corrections. These operations are widespread in human cognition. The chunks themselves serve as high level abstractions useful for explanation, communication, memory, and general reasoning, possibly forming the basis of the schemata and scripts of cognition.

Mechanisms

CSNNs provide a convenient computational architecture for this type of processing, but the IA model needs to be generalized in ways besides learning. The system should be able to process input strings of arbitrary length, and the network should be able to encode any number of levels of hierarchy. To accomplish both goals, I use weight sharing, a common practice in feedforward networks, but virtually never utilized for CSNNs. The entire network is shifted and duplicated to accommodate inputs wider than the longest chunks, and the smaller chunks are shifted and duplicated within the umbrella of the largest chunks, as shown in Figure 1.

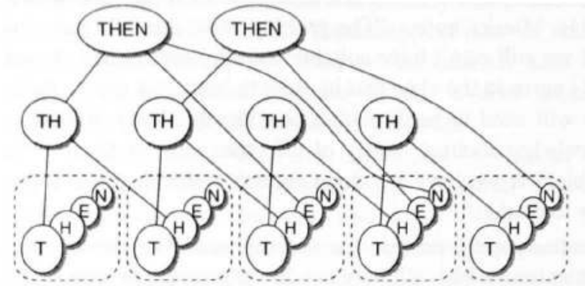

Figure 1: Part of a CSNN representing a 3-level CH with 4 and 2-letter chunks, applied to a 5-letter input.

As a first step for learning, I use a recently introduced algorithm called SEQUITUR (Nevill-Manning, 1996) to generate the CH encoded by the network. SEQUITUR incrementally builds a CH that can be used for examination or data compression, but has no mechanism for prediction. The marriage of this network formalism with an incremental CH generator provides a unique system capable of capturing wider and wider data interactions.

Preliminary experiments show that the qualitative processing behavior does generalize to more complex CHs and longer inputs, and that this structure can be learned. Weight tuning and a constructive CSNN chunking mechanism more sensitive to the statistical correlations in the data should provide better predictions and more compact hierarchies.

Novel combinations of previous ideas: weight sharing with CSNNs, a CH generation algorithm with CSNNs, and in the near future, a constructive network algorithm with CSNNs, provide a useful generalization of the successful IA model. Eventually, methods for discovering compositional structure will lead to synergistic benefits when combined with existing methods for discovering hierarchical taxonomic structure.

References

McClelland, J. L., & Rumelhart, D. E. (1983). An interactive activation model of context effects in letter perception: Part 1. an account of basic findings. *Psychological Review, 88*, 375–407.

Nevill-Manning, C. G. (1996). *Inferring Sequential Structure.* Doctoral Dissertation, Department of Computer Science, University of Waikato, New Zealand.