Unsupervised Learning of Conceptual Representations

- A Computational Neural Model

Michael Klein, Stefan Frank, Henk van Jaarsveld, Louis F. M. ten Bosch,

and Lou Boves

Levelt, Roelofs, and Meyer (1999) argue that concepts cannot be represented as feature clusters, but need to be represented as non-decompositional conceptual units. One major reason for this is the hypernym/hyponym problem: if concepts are represented as feature clusters, the conceptual representation of any word has at least all the features of its hypernym. It would therefore be impossible to activate, e.g., the word “mother” without triggering the word “parent” at the same time. However, if concepts are indeed represented as non-decompositional units, at least two other problems remain to be solved (given that the input to the semantic system are perceptual features): (i) explaining the emergence of such units during the acquisition of conceptual knowledge and (ii) preventing the analogue to the hypernym/hyponym problem to occur at the conceptual level, i.e., the problem that the input features leading to the activation of conceptual unit [[mother]] would also trigger the activation of [[parent]].

We present a computational neural model that solves both problems using a competitive-layer architecture. We ran a number of simulations using 77 different (binary) semantic features as input to the model during training and testing. Over 4900 training episodes, the features were presented in combinations that allowed the model to learn 49 concepts. These concepts could be grouped into 7 semantically related families. The maximum depth of the conceptual hierarchy was 3 (e.g., [[animal]], [[bird]], [[robin]]). We show that the model develops 49 conceptual units, each of which activates the strongest when the feature set that defines the concept is presented to the model. Therefore, our model can be regarded as a potential computational neural theory of how non-decompositional conceptual units emerge in the brain.

During the phase of concept acquisition, the model displays overgeneralizations (e.g., the use of the [[dog]] concept for everything with four legs) and undergeneralizations (e.g., the use the of [[dog]] concept for only the dog of the neighbours), which are in accord with data reported in the literature (e.g., Clark, 1973). After training, the presentation of an input vector of perceptual features always maximally activates the most precise concept: when the features of [[mother]] are presented, the [[mother]] concept is activated more strongly than the [[parent]] concept, and vice versa, despite the strong overlap in features. This shows that the conceptual analogue of the hypernym/hyponym problem does not occur.