A cognitive model that learns contextually appropriate representation scheme

松香 敏彦
千葉大学

key words : cognitive modeling, category learning, neural networks

Traditionally, it has been considered that there is a single innate internal representation system for categorical knowledge, such as Exemplars, Prototypes, or Rules. However, results of recent empirical and computational studies collectively suggest that the human internal representation system is apparently capable of exhibiting behaviors consistent with various types of internal representation schemes. We, then, hypothesized that humans’ representational system as a dynamic mechanism, capable of selecting a representation scheme that meets situational characteristics, including complexities of category structure. The present paper introduces a framework for a cognitive model that integrates robust and flexible internal representation machinery.

As in several other cognitive models of category learning, our model assumes that humans hold (memorized) internal reference points (i.e., rules, prototypes, or exemplars) and utilize psychological similarities or conformities between the input stimulus and the reference points as evidence to probabilistically assign the input stimulus to an appropriate category. In particular, psychological similarity (s_j) between input stimulus, x, and jth reference point, R_j is given as follows:

\[ s^*_j = \exp\left[-\beta \left(\sum_{n=1}^{\alpha} \alpha^*_n (R^*_j - x_i)^2 + \sum_{i=1}^{\tau} 2C^*_j R^*_j x_i (R^*_j - x_i)\right)\right] \]

\[ \alpha^*_n = \left[1 + \exp\left(-D^*_j\right)\right]^{-1} \]

where \( \beta \) is a constant that the experimenter can define in order to manipulate an overall similarity gradient. \( D_j \) and \( C_j \) are \( B_j \)'s dimensional and correlation selective attention, respectively. Superscript \( n \) is an index for different notions, and Subscripts \( i \) and \( m \) indicate feature dimensions, and \( I \) is the number of feature dimensions. Note that it is assumed that \( C_{jm} = C_{jm}, R^*_j \leq \alpha_j \alpha_m \leq \alpha_j \alpha_m \).

The information on similarities is then used to calculate category activations

\[ O^*_c(x) = \sum_j w^*_j s^*_j \]

where \( w^*_j \) is learnable association weights between exemplar \( j \) and category node \( k \), representing how strongly or weakly category nodes and reference points are coupled.

SUPERSET explicitly assumes that human learning is an optimization of a subjectively and contextually defined utility of knowledge being acquired. For example, like ordinary humans, SUPERSET can learn to acquire manageably simple yet sufficiently accurate knowledge. In other words, although SUPERSET incorporates rather complex internal representation mechanisms, it needs not to acquire “unnecessarily” complex concept. In fact SUPERSET can behave like several traditional models of category leaning as shown below.

**Exemplar-like representation:** When SUPERSET learns to pay no attention to feature correlations (i.e., \( C_{jm} = 0 \)) and has an identical dimensional attention for all exemplars (i.e., \( D_j = D_\text{im} \), \( \forall j \& k \)), then it would behave like a traditional exemplar model. There is no restriction on its association weights.

**Prototype-like representation:** When SUPERSET identifies and uses one prototypical exemplar for each category, then it behaves as a type of prototype model. To use only a few prototypical exemplars for each category, SUPERSET needs to deactivate several other remaining exemplars by weakening links between the remaining exemplars and the category nodes, i.e., \( w^*_j \leq \zeta, \forall j \notin \text{prototype} \), where \( \zeta \) is some threshold value. There is no restriction on its attention operations.

**Rule-like representation:** When a) SUPERSET learns to pay no attention to feature correlations (i.e., \( C_{jm} = 0 \)), b) the absolute value of selective dimensional attention to diagnostic feature dimensions are sufficiently large for all applicable rules (i.e., \( D_j > \zeta, \forall j \& i \notin \text{rule} \), where \( \zeta \) is some threshold value.), and c) only applicable rules have significant association with category nodes (i.e., \( w^*_j \leq \zeta, \forall j \notin \text{rule} \)), then SUPERSET behaves like a rule model.

**Hybrid representation:** Any other coefficient configuration can be interpreted as a hybrid representation scheme. SUPERSET can acquire, for example, a “rule plus exceptional exemplars” or a “prototype plus exceptions” like representation.

Three simulation studies were conducted. The results showed that SUPERSET successfully exhibited cognitive behaviors that are consistent with three main theories of the human internal representation system. Furthermore, a simulation study on social cognitive behaviors showed that the model was capable acquiring knowledge with high commonality, even for a category structure with numerous valid conceptualizations.