Allocation of Attention in Neural Network Models of Categorization

Toshihiko Matsuka (tm249@columbia.edu)

James E. Corter (jec34@columbia.edu)

Department of Human Development, Teachers College, Columbia University, 525 W. 120th St., New York, NY 10027 USA

Arthur B. Markman (markman@psy.utexas.edu)

Department of Psychology, University of Texas, Austin, TX 78712 USA

We compared ALCOVE (Kruschke, 1992), RASHNL (Kruschke & Johansen, 1999), SUSTAIN (Love & Medin, 1998), and the Cortico-Hippocampal Model (CHM) (Gluck & Myers, 1993) to see how they account for selective attention in category learning. Such comparisons may usefully augment comparisons of the models' classification accuracy.

Method

We simulated the results of studies of classification learning by Medin and Schaffer (1978) and Medin, Altom, Edelson & Freko (1982). The parameter values used for each model were adjusted to minimize the SSE in reproducing the training classification responses by human subjects.

Attention allocation predictions for the models were derived as follows. ALCOVE and RASHNL have explicit attention weight parameters, which are reported below. For SUSTAIN, the dimension-specific tuning parameters, λ, are reported. In the CHM there are no explicitly defined dimension attention parameters. We defined implicit measures of a dimension's attentional salience, by summing the absolute values of weights from all input nodes associated with a given dimension to the hidden node layer in the hippocampal net component of the CHM. To enhance comparability among the models, we computed and report relative attention weights for all the models.

Summary of Results

For Experiment 2 of Medin and Schaffer (1978), all the models fit the training set classification probabilities roughly equally well, but RASHNL and SUSTAIN were somewhat more accurate in predicting classification responses for the transfer stimuli. In this stimulus structure Dimensions 1 and 3 are highly predictive of the binary classification task, and Dimension 4 is moderately predictive. Somewhat surprisingly, ALCOVE, RASHNL, and the CHM gave as much or more attention weight to Dimension 4 as to the more diagnostic dimensions.

For Experiment 4 of Medin, Altom, Edelson & Freko (1982), RASHNL and the CHM fit the training set classification probabilities best, but RASHNL was the best and the CHM worst in predicting the transfer classifications. In this stimulus structure, Dimensions 1 and 2 are diagnostic in the sense that each is highly correlated with the criterion classification response, but Dimensions 3 and 4 have a simple XOR pattern in regards to the criterion classification. ALCOVE, RASHNL, and SUSTAIN all learn to allocate more attention to Dimensions 3 and 4, that together define the classification in terms of a simple XOR relationship. In contrast, the CHM pays more attention to the individually, but merely probabilistically, diagnostic Dimensions 1 and 2.

Conclusions

The four models give different predictions about attention weights for some stimulus structures. Examining and comparing these predictions may shed light on how the models learn. A promising line for future research is to gather direct data on how humans allocate attention in category learning (Matsuka, 2002).

References

Gluck, M. A., & Myers, C. E. (1993). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus, 3,* 491-516.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22-44.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 1083-1119.

Love, B. C., & Medin, D. L. (1998). SUSTAIN: A model of human category learning. *Proceeding of the Fifteenth National Conference on AI (AAAI-98),* 671-676.

Matsuka, T. (2002). Attention processes in category learning. Unpublished doctoral dissertation (draft), Teachers College, Columbia University.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8,* 37-50.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207-238