Title
Stochastic learning in neural network models of categorization

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 25(25)

ISSN
1069-7977

Authors
Matsuka, Toshihiko
Corter, James E.

Publication Date
2003

Peer reviewed

Stochastic learning in neural network models of categorization

Toshihiko Matsuka (matsuka@psychology.rutgers.edu)
Department of Psychology, Rutgers University - Newark
327 Smith Hall, 101 Warren St., Newark, NJ 07102 USA

James E. Corter (jec34@columbia.edu)
Department of Human Development, Teachers College, Columbia University
525 West 120th St. New York, NY 10027 USA

A number of neural network models of categorization have been proposed. The models differ notably in the types of internal representation used (e.g. exemplars vs. prototypes; hyperplane vs. hypersphere activation regions). However, many of these NN models of categorization (e.g., ALCOVE) use some form of gradient method for learning. These methods have been successful in reproducing group learning curves, but tend to underpredict variability in individual-level data, for both accuracy and attention allocation measures (Matsuka, 2002).

Here, we show that use of a different learning algorithm with a given model can result in different learning trajectories and more realistic variability in individual learning curves, especially for attention allocation. Our proposed algorithm is a form of constrained simulated annealing (Ingber, 1989). Initial parameter sets (dimensional attention weights and network connection weights) are randomly selected. At the beginning of each training epoch, a hypothetical "move" in the parameter space is computed by adjusting each parameter by an independently sampled term. These adjustment terms are drawn from a prespecified distribution (e.g., a Cauchy distribution). The move (i.e., the set of new parameter values) are accepted or rejected, based on the computed relative fit of the new values. Specifically, if the new parameter values result in better fit, they are accepted. If they result in worse fit, they are accepted with some probability *P*. The adjustment in parameters is very rapid initially, and it gradually decreases over learning blocks.

Simulations

In a simulation study, we modified ALCOVE (Kruschke, 1992) to incorporate stochastic learning. We compared ALCOVE with stochastic learning (ALCOVE-SL) to standard ALCOVE to compare how the models account for individual differences in category learning. To do this, we simulated the results of a classification learning study by Matsuka (2002). In this study, there were two perfectly redundant feature dimensions, Dimension 1 & Dimension 2. Besides classification accuracy, data on the amount of attention allocated to each feature dimension was collected. Results indicated that many subjects allocated attention to one or the other of the correlated dimensions, but not to both.

Both ALCOVE and ALCOVE-SL were able to reproduce the observed group learning curve. However, the amounts of attention allocated to Dimension 1 and 2 were identical for ALCOVE, while those for ALCOVE-SL tended to be distributed unequally. For comparison, we also tried using random initial values for ALCOVE's attention allocation parameters. In this version of ALCOVE, the amount of attention allocated to Dimension 1 and 2 were unequal initially, but showed virtually identical learning curves. Thus the stochastic-learning version of the model provided the best simulation of the empirical attention allocation curves.

Discussion

Our main goal in exploring new learning algorithms is to give NN models of categorization the capability to account for individual differences in distribution of attention. The simulation studies showed that the new algorithm is satisfactory in this regard. Stochastic learning algorithms have other desirable properties as well. For example, they can result in very rapid shift in attention allocation, which has been observed in human data (e.g., Kruschke & Johansen, 1999; Macho, 1997). Also, it could be argued that stochastic learning may be more psychologically plausible than gradient-based methods, which require more mental effort and assume that optimal adjustments are made to the vector of parameters on each trial.

References

Ingber, L. (1989). Very fast simulated annealing. *Journal of Mathematical Modeling, 12*: 967-973

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

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

Macho, S. (1997). Effect of relevance shifts in category acquisition: A test of neural networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 30-53.

Matsuka, T. (2002). Attention Processes in computational models of category learning. Unpublished doctoral dissertation. Columbia University, New York, NY.