In Proceedings of the Sixth International Conference on Cognitive Modeling, 196-201.
Mahwah, NJ: Lawrence Earlbaum.

Modeling Category Learning with Stochastic Optimization Methods

Toshihiko Matsuka (matsuka@psychology.rutgers.edu)
RUMBA, Rutgers University - Newark

James E. Corter (corter@tc.exchange.columbia.edu)
Department of Human Development, Teachers College, Columbia University

Abstract

Many neural network (NN) models of categorization (e.g., ALCOVE) use a gradient algorithm for learning. These methods have been successful in reproducing group learning curves, but tend to underpredict variability in individual-level data, particularly for attention allocation measures (Matsuka, 2002). In addition, many recent models of categorization have been criticized for not being able to replicate rapid changes in categorization accuracies and attention processes observed in the empirical studies (Macho 1997; Rehder & Hoffman, 2003). In this paper we introduce stochastic learning algorithms for NN models of human category learning and show that use of the algorithms can result in (a) rapid changes in accuracies and attention allocation, and (b) different learning trajectories and more realistic variability in individual-level.

Introduction

Recent NN models of classification learning, for example ALCOVE (Kruschke, 1992), RASHNL (Kruschke & Johansen, 1999), and SUSTAIN (Love & Medin, 1998), share a number of common aspects, including multilayer architectures and learned dimensional attention weights as well as learned association weights between stimulus input nodes and the output layer. One of these common elements is the use of *gradient-based learning algorithms* to adjust both association weights and dimensional attention parameters. In this method, weights are adjusted based on discrepancies between a training signal and the output activations on the current layer, by computing the *gradient* of the error function in the multidimensional parameter space. This is accomplished by taking the partial derivative of the error function with respect to each of the network parameters (weights) in turn. The algorithm then adjusts each of these weights proportionally to its partial derivative. This learning method is an effective means of finding optimal estimates for parameters, as long as the overall error function is not characterized by strong local minima. Thus the algorithm has normative justification (i.e., it models how people "should" learn or process information).

But, is the gradient method plausible *descriptively* (i.e., does it describe how people actually learn)? It seems implausible that people explicitly compute the gradient on each trial while attempting a classification learning task. On the other hand, people's general learning mechanisms might have evolved so as to approximate gradient learning. But in any case, we should first ask if the gradient-based learning algorithms are successful in replicating empirical data in human category learning. Results in the literature demonstrate that these methods have been successful in reproducing group learning curves (e.g., Kruschke, 1992, Kruschke & Johansen, 1999; Love & Medin, 2000). However, recent studies in our lab suggest these models may underpredict variability in individual-level empirical data, particularly differences in attention allocation measures (Matsuka, 2002; Matsuka & Corter, 2003; Matsuka, Corter & Markman, 2003). In addition, some empirical studies suggest that human's attention allocation to individual dimensions can change quite rapidly (Macho, 1997; Rehder & Hoffman, 2003). Most cognitive models based on gradient-based learning mechanisms appear to have difficulty reproducing such rapid changes in attention. In the present research we explore alternative learning algorithms for NN models of classification learning, specifically stochastic learning algorithms based on simulated annealing.

Stochastic Learning

Our proposed algorithm is based on a specific simulated annealing algorithm (Ingber, 1989). In the present algorithm, initial association weights are randomly selected from a uniform distribution centered at 0, and initial dimension attention strengths are equally distributed across all dimensions. This equal attention allocation in the early stages of learning is motivated by the results of empirical studies (Matsuka, 2002; Rahder & Hoffman, 2003) that showed many subjects tended to evenly allocate attention to the feature dimensions initially. In our algorithm, 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 Cauchy distribution. The move (i.e., the set of new parameter values) is then 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*. This probability is a function of a parameter called the "temperature", which decreases across blocks according to the annealing schedule. This particular annealing algorithm is relatively efficient, in that the adjustment in the network parameters is very rapid initially, and gradually decreases over learning blocks.

Key Assumptions

This learning algorithm can be applied to any feed-forward NN model of category learning. However, we assume that there is no (back) propagation of classification

errors in the models. Rather, we propose a very simple operation (i.e., comparison of two values) along with the operation of stochastic processes as the key mechanisms in category learning.

I. Initial network association weights ($\boldsymbol{w}$) are set to small random values, and initial dimension attention weights ($\boldsymbol{\alpha}$) are set equal across dimensions.
II. In learning, the attention strengths and association weights are updated with a random move in the parameter space, based on a prespecified distribution (e.g., the Cauchy distribution).
III. If the new $\boldsymbol{\alpha}$ and $\boldsymbol{w}$ result in better categorization accuracies (based on a "mini-simulation" using the network model), then the current move is accepted, and the new attention strengths and association weights replace the old values. In the case of a *decrease* in categorization accuracy due to the move, the move is accepted with some probability P ($0<P<1$) that is functional of both magnitude of error and moment of training (i.e., temperature).
IV. P is relatively large in the early stages of learning, but it decreases as learning progresses. This decrease is associated with a decrease in a parameter called the "temperature", by analogy with the physical process that occurs as a metal cools.

Thus, the present model does not assume that learning is associated with monotonic increases in accuracy (and attention) or continuous search for better categorization processes by human. Rather, it models random fluctuations or "errors" in people's memory and learning processes, and how people utilize and "misutilize" such errors.

As a test of these ideas, we have embedded the present learning algorithm into the ALCOVE model (Kruschke, 1992).

ALCOVE

ALCOVE (Kruschke, 1992), for Attention Learning COVEring map, is an exemplar-based multi-layer adaptive network model of categorization based in part on the Generalized Context Model or GCM (Nosofsky, 1986). The first layer of ALCOVE is a stimulus input layer. Each dimension has an attention strength (α_i) associated with it. The next layer in the network is the exemplar layer. Each node in this layer corresponds to an exemplar, described by its position in the multidimensional stimulus space, and receives input from the input layer. The activation of each exemplar node is calculated based on its similarity to the presented stimulus:

$$h_j = \exp\left[-c\left(\sum_i \alpha_i \mid \psi_{ji} - x_i \mid\right)\right] \quad \text{(A1)}$$

where ψ_{ji} is the value of exemplar node j on dimension i, x_i is the activation of input feature dimension i, c is a constant called the *specificity* that controls overall attention, and α_i is the attention strength for dimension i. In ALCOVE, the attention strengths essentially stretch and shrink dimensions.

The activity of the exemplar nodes is fed forward to the third layer, the category layer, whose nodes correspond to the categories being learned. The strength of association between category node k and exemplar node j is denoted by w_{kj}. The activation of category node k is then computed as the sum of weighted activations of all exemplars, or

$$y_k = \sum_j w_{kj} h_j \quad \text{(A2)}$$

The probability that a particular stimulus is classified as category k, denoted as $P(K)$, is assumed equal to the activity of category k relative to the summed activations of all categories, where the activations are first transformed by the exponential function (Kruschke, 1992):

$$P(K) = \frac{\exp(\phi y_k)}{\sum_k \exp(\phi y_k)} \quad \text{(A3)}$$

where ϕ is a real-value mapping constant that controls decisiveness of classification responses.

The standard version of ALCOVE (Kruschke, 1992) uses a form of gradient descent for updating weights. The error term is defined as the sum of squared differences between the desired and the predicted outputs:

$$E = \tfrac{1}{2}\sum_k (t_k - y_k)^2 \quad \text{(A4)}$$

Partial derivatives of the error function with respect to the association weights w_{kj} and the attention strengths α_i are used to compute the weight update:

$$\Delta w_{kj} = \frac{\partial E}{\partial w_{kj}} = \lambda_w (t_k - y_k) h_j \quad \text{(A5)}$$

$$\Delta \alpha_i = \frac{\partial E}{\partial \alpha_i} = -\lambda_\alpha \sum_j \left[\sum_k (t_k - y_k) w_{kj}\right] h_j \cdot c \mid \psi_{ji} - x_i \mid \quad \text{(A6)}$$

where λ_w and λ_α are the learning rates for the association weights and attention strengths, respectively. It is this gradient-based learning method that we propose to replace with the stochastic learning method.

Stochastic Learning Algorithms

Here, we have evaluated two applications of stochastic learning to ALCOVE: one version in which we implement stochastic learning for adjusting both dimensional attention weights and the network association weights (ALCOVE-CSL, for "completely stochastic learning"), and one in which stochastic learning is used to adjust only the dimension attention weights in ALCOVE (ALCOVE-SAL, for "stochastic attention learning ").

ALCOVE-CSL algorithm

STEP 0: Initialize:
Problem specific parameters (T^0,υ)
T^0 : initial temperature.
υ : temperature decreasing rate

Association weights w_{kj},
w_{kj} ~ U(MIN_w, MAX_w). where MIN_w and MAX_w are minimum and maximum values for $\boldsymbol{w}$.

Attention strengths α_i ,
$\alpha_i = 1/I^*(MAX_\alpha - MIN_\alpha)+MIN_\alpha$, for all $i = 1\ldots I$, where I is the number of feature dimensions.

Exemplar ψ_{ji}

$\psi_{ji} = x_{*i}$, where subscript * indicates unique patterns.

STEP 1: Calculate output activations

$$O_k = \sum_J w_{kj} \exp\left[-c\left(\sum_I \alpha_i \mid \psi_{ji} - x_i \mid\right)\right] \quad \text{(S1)}$$

STEP 2: Calculate fit index for one training block:

$$f(w^\tau, \alpha^\tau) = \sum_{n=1}^{N}\sum_{k=1}^{K}(d_k - O_k)^2 \quad \text{(S2)}$$

where K = # categories, N = # input in one block, d_k is a desired output for category node k at time τ.

STEP 3: Accept all weight and attention parameters ($\boldsymbol{\alpha}$ & $\boldsymbol{w}$) at the probability of:

$$P(w^\tau, \alpha^\tau \mid w^s, \alpha^s) = \left\{1 + \exp\left(\frac{f(w^\tau, \alpha^\tau) - f(w^s, \alpha^s)}{T^\tau}\right)\right\}^{-1} \quad \text{(S3)}$$

if $f(w^\tau, \alpha^\tau) > f(w^s, \alpha^s)$, or 1 otherwise, where $f(w^s, \alpha^s)$ is the fit index for the previously accepted parameter set.

STEP 4: Reduce temperature:

$$T^t = T^o \exp(-\upsilon \cdot t) \quad \text{(S4)}$$

STEP 5: Generate new $\boldsymbol{w}$ and $\boldsymbol{\alpha}$

$$w_{kj}^\tau = w_{kj}^s + y(MAX(w) - MIN(w)) \quad \text{(S5)}$$

$$\alpha_i^\tau = \alpha_i^s + y(MAX(\alpha) - MIN(\alpha)) \quad \text{(S6)}$$

where

$$y = \text{sgn}(u - 0.5)T\left[\left(1 + 1/T\right)^{|2u-1|} - 1\right] \quad \text{(S7)}$$

Here, u indicates a random number drawn from the Uniform distribution. S7 draws random numbers from the Cauchy distribution (Ingber, 1989).

REPEAT STEPS 1~5 until stopping criterion is met

ALCOVE with SAL

Stochastic Attention Learning (SAL) incorporates both gradient and stochastic methods for learning. In particular, SAL updates its association weights using a gradient method (Equation A5), and attention strengths by the stochastic learning method (Equations S6 & S7).

Since SAL incorporates gradient learning for its association weights, the badness-of-fit index at time t is often less than that at time $t-1$, even with an "inappropriate" random movement in attention allocation. In other words, the present algorithm as described in the previous section would accept many useless moves for attention distribution, particularly in the early stages of learning. However, this seems both unnecessary and inefficient. Thus, we modified SAL to include a threshold parameter ζ, which controls for the probability of accepting new attention weight values, to make the model accept only moves that satisfy a prespecified criterion (i.e., above the threshold) for categorization accuracy.

Thus for SAL, Equation S3 becomes

$$P(\alpha^\tau \mid \alpha^s) = \left\{1 + \exp\left(\frac{f(\alpha^\tau) + \zeta - f(\alpha^s)}{T^\tau}\right)\right\}^{-1} \quad \text{(S3')}$$

if $\{f(\alpha^\tau) + \zeta\} > f(\alpha^s)$, or 1 otherwise, where

$$f(\alpha^\tau) = \sum_{n=1}^{N}\sum_{k=1}^{K}(d_k - O_k)^2 \cdot$$

Simulations

In order to evaluate the abilities of the stochastic learning algorithms to account for human data on classification learning, we conducted three simulation studies. In Simulation 1, we tested if the stochastic learning can replicate rapid changes in attention allocation in category learning for a single simulated subject. In Simulation 2, we simulated the results of a recent empirical study on classification learning (Matsuka, 2002; Matsuka & Corter, 2003) to see if the algorithm can reproduce individual differences in attention processes. In Simulation 3, we examined if the algorithms accurately reproduce aggregated learning curves.

A total of four different ALCOVE-based models are involved in the simulations reported below. The main comparison we are interested in is to compare the performance of standard ALCOVE with ALCOVE incorporating stochastic attention learning (ALCOVE-SAL), and ALCOVE incorporating completely stochastic learning (ALCOVE-CSL). However, for Simulation 2, we also investigate if individual differences could be otherwise accounted for within standard gradient-based ALCOVE. To do this, we also tried another way (besides stochastic learning) of handling random individual differences within the ALCOVE model, namely by randomly varying individual learning rates. This version of standard gradient-learning ALCOVE is referred to here as ALCOVE-RLR.

Simulation 1: Rapid attention shifts

In the present simulation study, we examined if stochastic learning algorithms can replicate rapid changes in attention allocation as observed in empirical studies (Macho 1997, Rehder & Hoffman, 2003). Here, we used the simplest stimulus structure (T1) of Shepard, Hovland and Jenkins' stimulus sets (1961). Table 1 shows schematic representation of the stimuli used in the present simulation (i.e., T1).

Table 1: Schematic representation of stimulus set used in Simulations 1 and 3.

	Stimulus			Category					
	D1	D2	D3	T1	T2	T3	T4	T5	T6
1	1	1	1	A	B	A	A	A	A
2	1	1	0	A	B	A	A	A	B
3	1	0	1	A	A	A	A	A	B
4	1	0	0	A	A	B	B	B	A
5	0	1	1	B	A	B	A	B	B
6	0	1	0	B	A	A	B	B	A
7	0	0	1	B	B	B	B	B	A
8	0	0	0	B	B	B	B	A	B

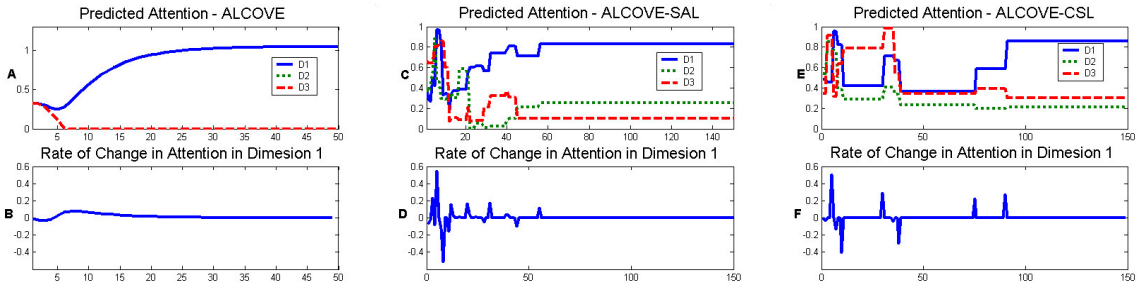

Figure 1: The results of Simulation 1. A: predicted attention allocation to the three feature dimension by standard ALCOVE; B: predicted rate of change in attention allocated to Dimension 1 by ALCOVE; C & E: predictions by ALCOVE-SAL; E & F by ALCOVE-CSL.

Simulation Method: Three ALCOVE-type models of category learning were evaluated in the present simulation studies, namely the standard ALCOVE, ALCOVE-SAL, and ALCOVE-CSL. They were run in a simulated training procedure to learn the correct classification responses. ALCOVE was run for 50 blocks of training, where each block consisted of a complete set of the training instances, while ALCOVE-SAL and ALCOVE-CSL were run for 150 blocks.

For each model, the gradient or rate of change in attention allocated to Dimension 1 was calculated by subtracting the amount of attention allocated to Dimension1 at time *t*-1 from that of time *t*. This measure was used as an index of how rapidly attention distributions changed.

Results: The results of one simulated subject for each model are shown in Figure 1. Note that only accepted attention strengths were graphed for ALCOVE with SAL and CSL. All three models learned to allocate the highest amount of attention to Dimension 1 and learned to ignore or pay less attention to Dimensions 2 and 3. The rate of attention change for ALCOVE was very smooth and its magnitude was smaller than those of ALCOVE-SAL and ALCOVE-CSL. ALCOVE-SAL and ALCOVE-CSL produced oscillating graphs with higher magnitudes of change. These results suggest that our proposed stochastic learning algorithms are capable of replicating rapid changes in attention allocation, while ALCOVE with gradient-based learning is not.

Simulation 2: Individual Differences

In this simulation study, we examined how the models account for individual differences in attention learning. To do this, we simulated the results of an empirical study on classification learning, Study 2 of Matsuka (2002). In this study, there were two perfectly redundant feature dimensions, Dimension 1 & Dimension 2 (see Table 2), and those two dimensions are also perfectly correlated with the category membership. Thus, information from only one of the two correlated dimensions was necessary for perfect categorization performance.

Besides classification accuracy, data on the amount of attention allocated to each feature dimension was collected in the empirical study. The measures of attention used were based on feature viewing time, as measured in a MouseLab-type interface (Bettman, Johnson, Luce & Payne 1993).

To summarize the empirical results that we are trying to simulate, 13 out of 14 subjects were able to categorize the stimuli almost perfectly (Figure 2, left top panel), and on average subjects paid attention to both of the correlated dimensions approximately equally (Figure 2, left middle panel). However, the examination of aggregated data can be misleading. When Matsuka and Corter (2003) analyzed attention data at an individual-level, they found that many subjects tended to pay attention primarily to only one of the two correlated dimensions, particularly in the late learning blocks (Figure 2, left bottom panel). This suggests that the participants utilized only the minimal necessary information for this task.

Table 2: Stimulus structure used in Study 2 of Matsuka (2002)

Category	Dim1	Dim2	Dim3	Dim4
A	1*	1*	3	4
A	1*	1*	4	1
A	1*	1*	1	2
B	2*	2*	2	1
B	2*	2*	3	2
B	2*	2*	4	3
C	3*	3*	1	3
C	3*	3*	2	4
C	3*	3*	3	1
D	4*	4*	4	2
D	4*	4*	2	3
D	4*	4*	1	4

Simulation method: The final parameter values used for each model were chosen by a simulated annealing method (Ingber, 1989; Matsuka, Corter & Markman, 2003) to minimize the objective function (i.e., sum of squared error) in reproducing the classification accuracies by human subjects. The four models were run in a simulated training procedure to learn the correct classification responses for the stimuli of the experiment. ALCOVE and ALCOVE-RLR were run for 48 blocks of training, where each block consisted of a complete set of the training instances, while ALCOVE-SAL and ALCOVE-CSL were run for 480 blocks. For each model, the final results are based on 50 replications

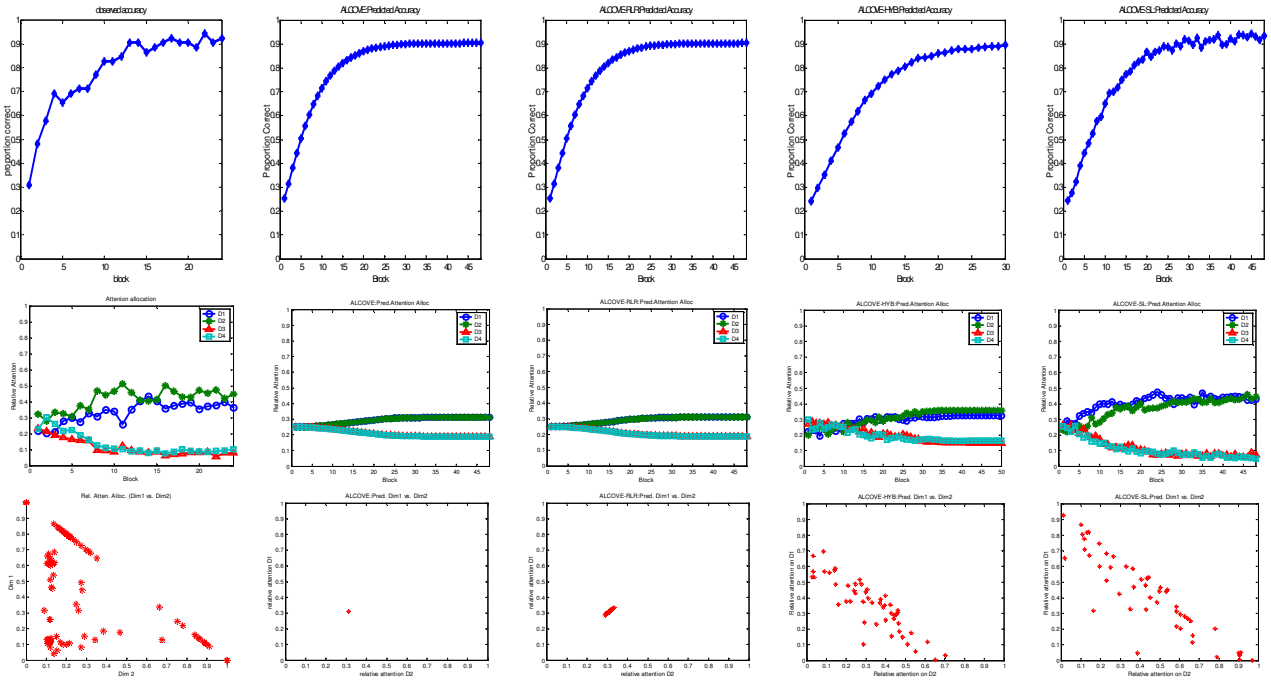

Figure 2: Results of Simulation 2. The left column shows observed learning curve (top panel), observed dimensional attention allocation (middle panel), observed attention allocation to Dimension 1 (Y-axis) and 2 (X-axis) in the last half of the training block (bottom panel). The predictions by ALCOVE, ALCOVE-RLR, ALCOVE-SAL, and ALCOVE-CSL are shown in the 2nd, 3rd, 4th and 5th column, respectively.

Results: Figure 2 summarizes the findings from the simulation study. The left column of this figure shows the empirical data from Study 2 of Matsuka (2002), including the learning curve for overall classification accuracy (top panel), the attention learning curves (middle panel), and attention allocated to Dimensions 1 and 2 (the redundant diagnostic dimensions) in the late training blocks (bottom panel).

Results for each of the models are shown in the remaining rows. Standard ALCOVE learns the classification task quickly, and allocates attention equally to the two diagnostic dimensions, but shows almost no inter-individual variability in attention, no intra-individual variability in the amount of attention allocated to Dimension 1 and 2 (Figure 2, Second row). ALCOVE-RLR, using random learning rate for attention strengths, but still using the standard gradient learning algorithm, shows some inter-individual variability, but virtually no intra-individual variability. ALCOVE-SAL, the version of ALCOVE modified to incorporate stochastic learning of attention weights, showed a less steep learning curve (like the human subjects), and more variability among subjects in attention allocation (again more closely resembling the empirical data). ALCOVE-CSL gave similar results, but exhibited some minor differences from ALCOVE-SAL (including more separation of the diagnostic and non-diagnostic dimensions). Note that the two versions of the stochastic learning models are able to replicate the observed learning curves.

Simulation 3. Replication of Nosofsky et al. 1994

So far, we have shown that our proposed stochastic learning algorithms are successful for reproducing individual-level data (i.e. rapid change & individual differences in attention processes). However, we merely tested the algorithms' capabilities of reproducing aggregated data. In the present simulation study, we simulated a classical study of categorization (Shepard et al. 1961) which is often used as a benchmarking stimulus set (e.g. Nosofsky, Gluck, Palmeri, McKinley & Glauthier, 1994). The stimulus structures are shown in Table 1. The results of previous empirical studies showed that Type 1 (T1) was the easiest to learn to classify, followed by T2, T3, T4, T5, and T6 being the most difficult. More precisely, Nosofsky et al. (1994) showed that the numbers of error made thus difficulties for those stimulus structures were significant except T3, T4, and T5.

Simulation method: In the present study, we tested only ALCOVE-CSL and ALCOVE-SAL, as the standard ALCOVE has previously been shown to be able to replicate the observed learning curves (Nosofsky et al. 1994). The two models were run in a simulated training procedure to learn the correct classification responses for the stimuli. ALCOVE-CSL was run for 250 blocks of training, where each block consisted of a complete set of the training instances, while ALCOVE-SAL was run for 150 blocks. For each model, the final results are based on 500 replications.

Results: Figure 3 summarizes the findings from the Simulation 3. Both ALCOVE-CSL and SAL were able to reproduce the order of difficulty successfully. That is ALCOVE-CSL and SAL find T1 to be the easiest, followed by T2, T3, T4, T5, and T6.

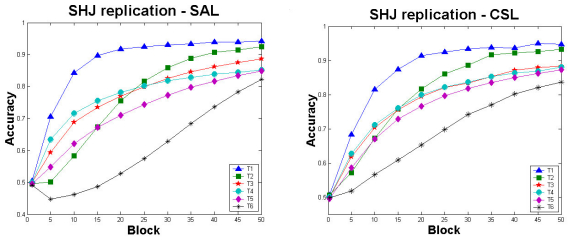

Figure 3. Results of Simulation 3. Both ALCOVE-SAL and ALCOVE-CSL were able to reproduce the order of difficulty successfully

Discussion and Conclusion

We have investigated the possibility of using stochastic learning rather than gradient-based methods in neural network models of human classification learning. In the present simulations we have explored the effectiveness of this method in several variants of the ALCOVE model (Kruschke, 1992). Our main goals were to see if stochastic learning algorithms 1) were able to replicate rapid change in attention processes, and 2) offered a better account of individual differences in final distribution of attention, particularly distribution of attention to the two perfectly correlated dimensions. The simulation studies showed that the new algorithms are satisfactory in these regards.

Stochastic learning algorithms have other desirable properties as well. 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. One caveat to these results is that the stochastic learning algorithms learn more slowly than the standard gradient methods in categorization tasks with relatively small (both number of stimulus feature dimensions and number of unique exemplars) and well-defined stimulus sets that are usually used in laboratory experiments. However, for more realistic category learning involving complex category structures and/or stimuli with many feature dimensions, stochastic learning may be able to learn faster than ordinary gradient type learning.

Distribution of random numbers: In the present research, the random moves in parameter space were drawn from the Cauchy distribution, mainly because its fatter tails are more likely than the Gaussian distribution to produce rapid and or large shift in attention allocation, which has been reported by some empirical studies. However, with a proper experimenter-defined parameter setting (e.g., initial temperature & temperature decreasing rate), similar results in attentional shifts might have been achieved with some other distributions, including normal, rectangular, skewed, or multi-modal distribution. In addition, we assumed the "temperature" decreases as a function of time or the number of training blocks. But, it may not capture realistic learning mechanisms. Rather, there may be more descriptively valid factor(s) for the annealing schedule, such as classification accuracy. Further simulation and empirical studies seem useful for investigating descriptive validity of stochastic optimization methods as a model of human category learning.

Acknowledgments

This study is partially supported by The National Science Foundation (the award EIA-0205178 granted to Stephan Jose Hanson).

References

Bettman, J.R., Johnson, E.J., Luce, M.F., Payne, J.W. (1993). Correlation, conflict, and Choice. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 931-951.

Ingber, L. (1998). Very fast simulated annealing. Journal of Mathematical Modelling, 12: 967-973.

Kruschke, J. E. (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, and 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.

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.

Matsuka, T. & Corter, J. E. (2003). Empirical studies on attention processes in category learning. Poster presented at 44th Annual Meeting of the Psychonomic Society. Vancouver, BC, Canada.

Matsuka, T., Corter, J. E. & Markman, A. B. (2003). Allocation of attention in neural network models of categorization. Under review

Nosofsky, R.M. (1986). Attention, similarity and the identification – categorization relationship. Journal of Experimental Psychology: General, 115, 39-57

Nosofsky, R.M., Gluck, M.A., Palmeri, T.J., McKinley, S.C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). Memory and Cognition, 22, 352-369.

Rehder, B. & Hoffman, A. B. (2003). Eyetracking and selective attention in category learning. Proceedings of the 25th Annual Meeting of the Cognitive Science Society, Boston, 2003.

Shepard, R.N., Hovland, C.L., & Jenkins, H.M. (1961). Learning and memorization of classification. Psychological Monograph, 75 (13).