In Proceedings of the Sixth International Conference on Cognitive Modeling, 190-195. Mahwah, NJ: Lawrence Earlbaum.

Exploratory Approach for Modeling Human Category Learning

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

Abstract

One problem in evaluating recent computational models of human category learning is that there is no standardized method for systematically comparing the models' assumptions or hypotheses. In the present study, a flexible general model (called GECLE) is introduced that can be used as a framework to systematically manipulate and compare the effects of a limited number of assumptions at a time. Two simulation studies are presented to show how the GECLE framework can be useful in the field of human high-order cognition research.

Introduction

The past fifteen years have seen significant advances in adaptive network models of categorization. In particular, three models of human category learning have attracted much attention, namely ALCOVE (Kruschke, 1992), RASHNL (Kruschke & Johansen, 1999), and SUSTAIN (Love & Medin, 1998). These models share many properties, because they can be considered special cases of Generalized Radial Basis Functions (Poggio & Girosi, 1990), as discussed by several authors (Kruschke, 1993; Matsuka & Corter 2003a; Rosseel, 1996). First, all three models are multilayer adaptive network models, with "reference points" ("basis units" in RBF terminology) in their memory (specifically, in the hidden layer). The models all use similarities between the reference points (RPs) and the input stimuli for calculating activations of RPs. Then, the weighted RP activations are fed forward to output nodes, whose activations are used to categorize the input stimuli. All three models scale feature dimensions independently in calculating these input-to-RP similarities, and this scaling process is interpreted as reflecting dimensional attention processes (Kruschke, 1993). In addition, all models incorporate as their basic learning method a form of gradient descent for incremental adjustments of both association weights and dimension-specific attention parameters. Finally, all three models may be considered as confirmatory models, because they are based on specific a priori assumptions about how humans process information in categorization (e.g. how stimuli are internally represented & how humans pay attention to stimuli). In order to test these specific assumptions, they should be varied systematically, preferably one at a time. The assumptions could then be tested by comparing the fit to empirical data of the resulting models. This is the major theme of the present work.

There are several differences among the three models as well. First, the assumptions about how stimuli are internally represented are different. ALCOVE and RASHNL are exemplar models, in the sense that each stimulus in the training set is allocated as a RP in the "hidden" layer of the networks, and the RPs reside in fixed locations. In contrast, SUSTAIN is a prototype model that uses a reduced number of movable RPs in its hidden layer, corresponding to potential generalizations. In addition, SUSTAIN dynamically allocates new prototypes, thus it may use multiple prototype nodes for each category explicitly defined by the training feedback. Second, how RP activations are utilized in making category predictions and in adjusting parameter estimates during learning are different for the models. SUSTAIN utilizes only the single most activated RP for categorization and learning, whereas ALCOVE and RASHNL utilize the activations of all RPs. Third, the assumptions about attention processes are different. RASHNL assumes limited attention capacity and rapid shifts in attention processes, whereas ALCOVE and SUSTAIN do not. Fourth, the functions for computing similarity measures and RP activations are different.

There have been several studies comparing computational models of categorization, including but not limited to ALCOVE, RASHNL, and SUSTAIN (e.g., Matsuka, Corter, & Markman, 2003; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994). Although these comparative studies provided information on the models' capabilities for reproducing human-like categorization learning, they did not necessarily provide information that can lead to specific understanding of the nature of human category learning. That is because model-to-model comparisons are not informative for testing the plausibility of each specific assumption, rather such model comparisons are essentially omnibus tests collectively comparing all variations in assumptions at once.

Since it has been difficult to use the results of these previous comparative studies to understand which specific assumptions are supported by the data, it seems desirable to have a general framework for modeling human category learning that allows us to manipulate and test one or a limited number of model assumptions at a time. The framework should be general and flexible, so that we can conduct standardized exploratory modeling of various types of human cognitive processes associated with categorization.

New Model of Human Category Learning

Qualitative Descriptions: The GECLE (for Generalized Exploratory models of Category LEarning) is a general and flexible exploratory modeling approach for human category learning, that is capable of modeling human category learning with many variants using different model assumptions. This general model allows model assumptions to be manipulated separately and independently. For example, one can manipulate assumptions about how stimuli are internally represented (e.g. exemplars vs.

prototypes), or about how people selectively pay attention to input feature dimensions (e.g., paying attention to dimensions independently or not).

The GECLE model uses the Mahalanobis distances (in quadratic form) between the internally represented reference points (corresponding to either exemplars or prototypes) and the input stimuli as the measure of similarity between them. Thus, unlike other NN models of category learning, the GECLE does not necessarily assume that attention is allocated independently dimension-by-dimension. Rather, it assumes that humans in some cases do pay attention to correlations among feature dimensions. This allows the GECLE to model processes interpretable as dimensionality reduction or mental rotation in the perception and learning of stimuli. Such processes may enhance the interpretability of stimuli for categorization task for humans. Another motivation for the use of Mahalanobis distance is that the capability for paying attention to correlations among feature dimensions may be needed for classification tasks defined on integral stimuli. In the GECLE framework, the attention parameters (which are the diagonal and off-diagonal elements of the covariance matrices) can be considered as *shape* and *orientation* parameters for receptive fields or attention coverage areas of the reference points. It should be noted, however, that one can constrain GECLE to incorporate the "dimensional attention processes" assumption (i.e., attention is allocated independently on a dimension-by-dimension basis) by forcing the off-diagonal entries in the covariance matrices to be equal to zero.

Another unique feature of GECLE's attention mechanism is that it allows each reference point to have uniquely shaped and oriented attention coverage area (Figures 1D, 1E, and 1F). I call this "local attention coverage structure". Again, one can impose a restriction on the model's attention mechanism by fixing all covariance matrices to be the same, which I refer as "global attention coverage structure" (Figures 1A, 1B, and 1C). The local attention coverage structure model is complex, but may plausibly model attention processes in human category learning. For example, it allows models to be sensitive to one particular feature dimension when the input stimulus is compared with a particular reference point that is highly associated with category X, while the same feature dimension receives little or no attention when compared with another reference point associated with category Y. Thus the local attention coverage structure causes models to learn and be sensitive to within-cluster or within-category feature configurations, while the global attention coverage structure essentially stretches or shrinks input feature dimensions in a consistent manner for all RP receptive fields and all categories.

Another way of interpreting GECLE's capabilities for paying attention to correlations among feature dimensions and having local attention coverage structures is that the model learns to define what the feature dimensions are for each RP and to allocate attention to those dimensions. In contrast, for almost all previous adaptive models of category learning, the definition of the feature dimensions is static and supplied by individuals who use the models.

The other notable characteristic of the GECLE's attention mechanism is that the user can manipulate characteristics of the distributions assumed for the activations of the reference points. For example, one can have RP activation distributions with thicker tails to obtain more competition among the RPs. In its natural form, the GECLE may be considered as a model using prototype internal representation, because it tries to learn to locate its reference points at centers of each category cluster. However, with a proper user-defined parameter setting, it can behave like a model with an exemplar-based internal representation.

Quantitative Descriptions (Algorithm): The feedforward and learning algorithms of the GECLE are typical for implementation of the Generalized Radial Basis Function (Haykin, 1999; Poggio & Girosi, 1990). GECLE uses the following function to calculate the distances or similarity between internally represented reference points (e.g., prototypes or exemplars) and input stimuli at time n:

$$D_j^n\left(x^n, r_j\right) = \left(x^n - r_j\right)^T \Sigma_j^{-1}\left(x^n - r_j\right) \tag{1}$$

where x^n is an I-ruple vector representing an input stimulus consisted of I feature dimensions presented at time n, r_j, also an I-ruple vector, that corresponds to the centroids of reference point j, expressing its characteristics, and Σ_j^{-1} is the inverse of the covariance matrix, which defines the shape and orientation of the attention coverage area of reference point j. For a model with global attention coverage structure, there is only one global Σ^{1} for all reference points. The entries (s_{im}) in Σ_j must satisfy the following conditions: $s_{ii} \geq 0$ & $|s_{im}| \leq$ MIN(s_{ii}, s_{mm}).

The psychological similarity measures $D_j(x,r)$ cause some activations in internal "hidden" units or reference points (i.e., exemplars or prototypes). The activation of "hidden" basis unit j, or h_j, is obtained by any differentiable nonlinear activation transfer function (ATF), or

$$h_j = G\left(D_j(x,r)\right) \tag{2}$$

given that its first derivative $G'(\cdot)$ exists. An exponential function, $\exp(-cD_j(x,r))$, is an example of an ATF. The ATF must be a differentiable function, because GECLE uses a gradient method for learning, where the partial derivatives are used for updating the learnable parameters. However, it is possible to eliminate this restriction by incorporating a form of derivative-free learning algorithm such as stochastic learning (Matsuka & Corter, 2004).

The activations of hidden units are then fed forward to output nodes. The activation of the kth output node, O_k, is calculated by summing the weighted activations of all hidden units connected to the output node, or

$$O_k = \sum_{j=1}^{J} w_{kj} h_j \tag{3}$$

where w_{kj} is the association weight between output node k and reference point j. The probability that a particular stimulus is classified as category C_k, denoted as $P(C_k)$, 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)

$$P(C_k) = \frac{\exp(\phi O_c)}{\sum_k \exp(\phi O_k)} \tag{4}$$

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

GECLE uses the gradient method to update parameters. The error function is defined as the sum of squared differences between targeted and predicted output values (i.e., L_2 norm), or

$$E(w,r,\Sigma^{-1}) = \frac{1}{2}\sum_{k=1}^{K} e_k^2 = \frac{1}{2}\sum_{k=1}^{K}\left(d_k - O_k\right)^2 \quad (5)$$

Then the following functions are used to update parameters.

$$\Delta w_{jk} = \frac{\partial E}{\partial w_{jk}} = -\eta^{w} e_k h_j \quad (6)$$

where η^{w} is the learning rate for the association weights.

$$\Delta r_j = \frac{\partial E}{\partial r_j} = -\eta^{r} \sum_{k=1}^{K} e_k w_{jk} G'\left(D_j(x,r)\right)\Sigma_j^{-1}\left(x - r_j\right) \quad (7)$$

where $G'(\cdot)$ is a derivative of $G(\cdot)$. Equation 7 can be considered as a function that locates or defines prototypes of stimuli. For the exemplar-based modeling η^{r} must be set to zero to maintain the static nature of reference points.

$$\Delta\Sigma_j^{-1} = \frac{\partial E}{\partial \Sigma_j} = \eta^{\Sigma} \sum_{k=1}^{K} e_k w_{jk} G'\left(D_j(x,r)\right)\left(x - r_j\right)\left(x - r_j\right)^{T} \quad (8)$$

For models with global attention coverage structure, Equation 8 should be summed over both k and j.

Varieties of Activation Transfer Function: The ATF in the GECLE can be any function as long as it is differentiable. This allows one to manipulate and compare the effects of specific characteristics of the population attention structure (e.g. fatter tail vs. thinner tail). This capability was motivated by the fact that the population attention structure can determine the effectiveness of model predictions. For example, Hanson and Gluck (1991) compared RBFs with Gaussian and Cauchy activation functions, and showed that increased competition by the Cauchy's fatter tails resulted in better fit to the empirical data. Since there is not enough evidence indicating the "true" or best activation transfer function, and to enhance the flexibility of GECLE, ATF is deliberately made user-definable.

Hierarchy of Constraints on Attention Parameters (Σ): There is a hierarchy of constraints that one can impose on the attention parameters Σ to manipulate GECLE's attention mechanisms. There are two levels of uniqueness of Σ (global and local attention coverage structure), in each of which there are three levels of constraints on entries in Σ. The following is a list of six possible levels of restriction, and Figure 1 shows examples of the corresponding attention coverage structures. Note that regardless of the types of restriction, the entries (s_{im}) in Σ_j must satisfy the following conditions: $s_{ii} \geq 0$ & $|s_{im}| \leq$ MIN(s_{ii}, s_{mm}).

Global Attention Coverage Structures

A. Global Pure Radial (GPR): Constraints on Σ_j: $s_{ii} = s$, for all i: $s_{im} = 0$, for all $i \neq m$; $\Sigma_j = \Sigma$, for all reference points j.

B. Global Uncorrelated Non-radial (GUN): Constraints on Σ_j: $s_{im} = 0$, for all $i \neq m$; $\Sigma_j = \Sigma$, for all reference points j.

C. Global Correlated Non-radial (GCN): Constraints on Σ_j: $\Sigma_j = \Sigma$, for all reference points j.

Local Attention Coverage Structures

D. Local Pure Radial (LPR): Constraints on Σ_j: $s_{ii} = s$, for all i; $s_{im} = 0$, for all $i \neq m$.

E. Local Uncorrelated Non-radial (LUN): Constraints on Σ_j: $s_{im} = 0$, for all $i \neq m$.

F. Local Correlated Non-radial (LCN): Constraints on Σ_j: none.

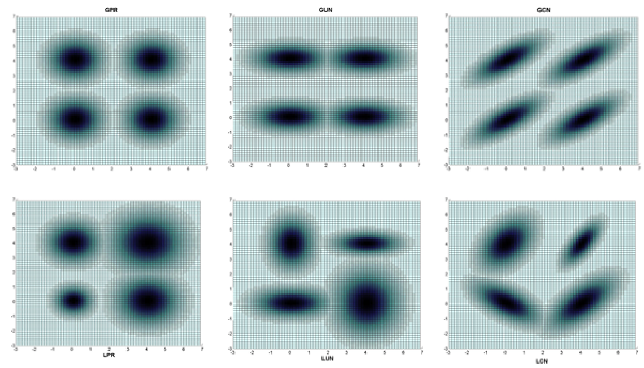

Figure 1. Six types of attention structures in the GECLE framework. Clockwise from top left panel, GPR, GUN, GCN LCN, LUN, and LPR.

Application s of GECLE

Comparing Internal Representation assumptions: There has been an increasing number of studies investigating and debating how stimuli are internally represented in human cognition during the last several years (e.g., Minda & Smith 2002; Nosofsky & Zaki 2002). Most of these debates have been based on quantitative models of categorization, and only a few have considered representational aspects of adaptive or network models of category learning. One limitation of the models of categorization is that, as Shanks (1991) pointed out, its model fitting process is post hoc and thus does not generate predictions on learning processes.

Several studies (Matsuka, 2002; Matsuka et al, 2003) have compared exemplar-based (EB) and prototype-based (PB) adaptive network models of category learning, but there have been no systematic comparison of specific assumptions in EB and PB modeling. For example, the EB and PB models compared in those studies assume different attention processes and utilize reference points differently for categorizing and learning. Thus, differences in the accuracy of reproducing learning curves may not be attributed solely to the plausibility of the EB versus PB representations, but possibly to multiple factors including the models' attention mechanisms. With GECLE, one can systematically compare the plausibility of the EB and PB representations by holding the attention mechanisms constant for both models (i.e., using the same activation transfer function and the same constraints on Σ^{1}).

Comparing Selective Attention Mechanisms: Selective attention processes have been suggested to play a very important role in human category learning (Shepard, Hovland, Jenkins, 1962). However, only limited numbers of selective attention mechanisms have been modeled and tested. For example, virtually all recent network models of categorization assume dimensional attention processes (i.e., no attention to correlations) and global attention coverage structure (i.e., all reference points have exactly the same shape of receptive field, with independent attention allocation to dimensions).

Again, a general framework like GECLE allows systematic manipulation of models' attention mechanism, and or exploration of various types of attention mechanism that has not be tested, such as distribution of attention to correlated features dimensions. In addition, comparisons on different activation transfer functions $G(\cdot)$ can be informative for understanding how human categorize stimuli.

Investigating Interactions Between Internal Representation & Attention Mechanisms. As a final point, it may be possible that a model with a particular internal representation system (e.g., exemplar-based) performs better with a particular attention mechanism, that does not work as well for models with other representation system (e.g., prototypes). In other words, the effectiveness of internal representation system and attention mechanism may interact with each other in the sense that the effectiveness of model's internal representation systems may depend on its attention mechanism, or the effectiveness of the models' attention mechanism may depend on its internal representation system.

I am not aware of any single study that systematically and simultaneously manipulates models' internal representation system and attention mechanism to investigate possible interaction effects between them. GECLE provides a way to systematically manipulate, by a factorial design, both internal representation system and attention mechanism to tackle this issue of interactivity of internal representation and attention mechanisms.

Simulations

In this section, simulation studies are conducted as examples showing how GECLE can be informative in the field of human category learning. In particular, two simulation studies, investigating possible interactive effects of the models' internal representation and selective attention mechanisms are reported.

Simulation 1: XOR problem

In Simulation 1, a simple exclusive-or (XOR) learning task is simulated with both prototype- and exemplar-based GECLE models. An XOR is one of the simplest stimuli structures with which one can expect interactions between representation system and attention mechanism.

There are eight different models involved in this simulation, namely, E1: an exemplar-based (EB) model with GUN attention mechanism; E2: EB with GCN; E3: EB with LUN; E4: EB with LCN; P1: a prototype-based (PB) model with GUN; P2: PB with GCN; P3: PB with LUN; and P4: PB with LCN. All EB models had four reference points, while all PB models had two. For all eight models, the following one-parameter exponential activation transfer function was used:

$$h_j = \exp\left(-c \cdot D_j(x,r)\right)$$

The models were run in a simulated training procedure to learn the correct classification responses for the stimulus set. The models were run for 250 blocks of training, where each block consisted of a complete set of the training instances. The user-defined parameters (e.g. learning rates) were selected arbitrary.

Results: Tables 1a and 1b show the results of Simulation 1. All exemplar-based models were able to learn to categorize XOR stimuli by utilizing four exemplars (i.e., all unique stimulus configurations) in their "memory", and thus complex attention mechanisms were shown to be ineffective or unnecessary for EB modeling. In fact, when the fit measure was adjusted for model complexity (i.e., number of parameters), the EB model with the simplest attention mechanism resulted in the best (relative) fit. In contrast, for prototype-based modeling, only LCN (i.e., P4) was able to learn to categorize the XOR stimulus set, suggesting that the complex attention mechanism plays an important role for the PB modeling (Figure 2 shows activation areas produced by P4). Note that P4 also resulted in the best (relative) fit among all eight models after controlling for its complexity.

In sum, the results of the present simulation suggest that it is very likely that effectiveness of the model's attention mechanism depends on how the stimuli are internally represented by the model or vice versa; a simple GUN attention mechanism seems sufficient for EB modeling, while a complex LCN is required for PB modeling.

Table 1a: Results of prototype-based GECLE

Model	P1	P2	P3	P4
Attention structure	GUN	GCN	LUN	LCN
No. prototypes	2	2	2	2
No. Learnable parameters	10	11	12	14
SSE	1.328	1.147	1.322	$\varepsilon^{\dagger}$

$\varepsilon^{\dagger}$ < 10e-20.

Table 1b: Results of exemplar-based GECLE

Model	E1	E2	E3	E4
Attention structure	GUN	GCN	LUN	LCN
No. Exemplars	4	4	4	4
No. Learnable parameters	18*	19*	24*	28*
SSE	$\varepsilon^{\dagger}$	$\varepsilon^{\dagger}$	$\varepsilon^{\dagger}$	$\varepsilon^{\dagger}$

$\varepsilon^{\dagger}$ < 10e-20.

* Location parameters for exemplars were static & not subject to error-minimization learning, but it is assumed that optimized locations are learned when the exemplars are created.

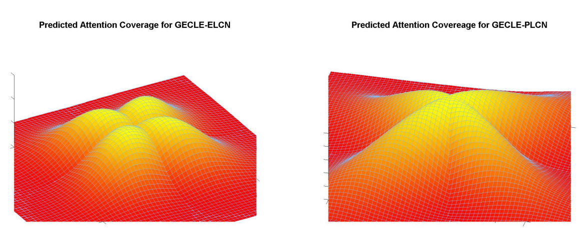

Figure 2: Activation areas & strength of E4 and P4

Simulation 2: Filtration vs. Condensation

Kruschke (1993) claimed that selective dimensional attention processes (i.e., paying attention to each dimension independently) is one of three key principles for models of category learning. His claim was based partly on the evidence that humans learn much better in “filtration” tasks, in which information from only one dimension is required for perfect categorization, than in “condensation” tasks, in which information from two (or more) dimensions is required (Gottwald & Garner, 1975; Kruschke 1993). Thus, a model paying attention to correlations or having diagonal attention coverage may not be able to show the filtration advantage, implying that any model with a GCN or LCN attention mechanism may not be able to replicate such an advantage. If the claim valid, then it is evidence against P4, namely the prototype-based model with diagonal localized attention coverage, as a descriptive model of human cognition. The present simulation study tests if PB model with LCN attention mechanism can replicate the filtration advantage observed in human category learning.

Method: In simulation 2, I revisited Kruschke’s claim regarding dimensional attention processes by simulating category learning on both filtration and condensation stimuli using the prototype-based model with LCN (and EB-LCN for a illustrative comparison). The stimulus set presented in Kruschke (1993) is used in this simulation. Table 2 shows the schematic representation of the stimulus set. For the filtration stimulus set, information from only Dimension 1 is required for a perfect categorization (category = A, if D1 < 2; category = B, otherwise) while information on both Dimensions 1 and 2 were required for the condensation set. The same one-parameter exponential ATF used in Simulation 1 is used in the present simulation study. The user-defined parameters were optimized using a simulated annealing method (Ingber, 1989; Matsuka et al. 2003) to reproduce observed empirical learning curves reported in Kruschke (1993).

It should be noted that Kruschke (1993) showed that ALCOVE (i.e., an EB-model with GUN) was able to reproduce the filtration advantage.

Table 2: Stimulus sets used in Simulation 2

Stimulus feature		Category	
Dim 1	Dim 2	Filtration	Condensation
0	1	A	A
0	2	A	A
1	0	A	A
1	3	A	B
2	0	B	A
2	3	B	B
3	1	B	B
4	2	B	B

Results: Figure 3 shows the results of Simulation 2. The prototype-based LCN model was able to show the filtration advantage even when it paid attention to the correlation between the two input dimensions. In contrast, the exemplar-based LCN model showed no filtration advantage.

One possible reason why PB-LCN showed the filtration advantages was that PB-LCN might have been able to locate or define prototypes more easily in the filtration task than in the condensation task. This is because while the condensation stimuli require synchronization of the “correct” movements of centroids of prototypes and the “correct” psychological scaling of the two feature dimensions (i.e., attention processes), the filtration stimuli require “correct” movements and scaling in only one dimension. Thus, synchronization of prototype-movement and scaling was more difficult for the condensation stimuli than in the filtration task for models using prototypes. In other words, category learning by any prototype-based network models is strongly affected by how successfully or how fast the models can find “proper” prototypes and how well psychological scaling of dimensions is synchronized with it. For E4, this was not a problem as it had exemplars in the correct locations from the beginning, resulting in no filtration advantage. These results indicate that correlated (i.e., diagonally-oriented) attention coverage may be more often a required assumption for PB modeling, compared to EB modeling.

As in Simulation 1, the results of the present simulation suggest that it is highly possible that stimuli’s internal representation and selective attention mechanisms interact with each other. Furthermore, our simulation studies suggest that human may allocate attention not only to individual dimensions, but also to correlations among dimensions. At the very least, the evidence of a filtration advantage observed in human subjects does not rule out the possibility that humans pay attention to correlations among feature dimensions.

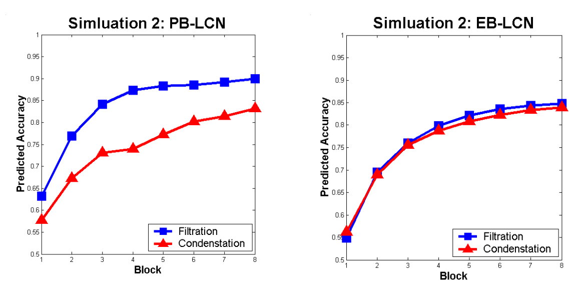

Figure 3: The results of Simulation 2

Discussion and Conclusion

Individual Differences. The results of some simulation studies (e.g., Matsuka, 2002, Matsuka & Corter, 2003b) suggest that NN models of categorization with gradient learning methods are successful in reproducing group learning curves, but tend to underpredict variability in individual-level data. Since GECLE utilizes a gradient method for learning, it too is expected to underpredict individual differences. However, this exploratory model is introduced to compare how well models with different representational and processing assumptions can replicate general tendencies in human category learning. Many of these general tendencies may be best described in terms of such aggregated data. Nonetheless, to account for individual differences, GECLE could be easily modified to incorporate Matsuka & Corter's (2004) stochastic learning algorithm for attention processes, which is shown to be more successful in reproducing individual differences in learning.

Conclusion

One of the most critical problems in evaluating recent computational models of categorization is that there is no standardized method for comparing the models' assumptions systematically. Thus, previous studies involving model comparisons have sometimes been unable to answer which element, assumption, or structure of each model was responsible for successful or unsuccessful replication of observed tendencies in human category learning. In the present study, a flexible general model is introduced, that can be used as a framework to systematically compare a limited number of assumptions at a time.

Two simulation studies are described to show how the GECLE framework can be useful in exploring issues in the field of categorization research. The results of Simulation 1 showed that a pure prototype-based category learning model (i.e., the number of prototypes is equally to that of category) was capable of learning an XOR problem only if it incorporated a very complex attention mechanism, while the exemplar-based model was capable of learning the stimuli with a simple attention mechanism. In Simulation 2, the filtration advantage, which has been used as an argument or evidence for dimensional attention processes (i.e., paying attention to dimension independently with no attention to correlations among feature dimensions), was successfully replicated by the prototype model with a complex attention mechanism capable of paying attention to correlation. This casts some doubt on the claim that the filtration advantage shows that people pay attention to dimensions independently, without attending to correlations among dimensions.

The results of these simulations are argued to provide new insights regarding human category learning, namely that 1) it is very likely that there are interactions between internal mental representation and attention mechanisms, and 2) people may pay attention to correlations among feature dimensions.

References

Gottwald, R. L. & Garner, W. R. (1975). Filtering and condensation tasks with integral and separable dimensions. *Perception & Psychophysics, 2*, 50-55.

Hanson, S. J., & Gluck, M. A. (1991). Spherical units as dynamic consequential regions: Implications for attention and cue-competition in categorization. *Advances in Neural Information Processing Systems #3*. San Mateo, CA: Morgan Kaufman, 656-665.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation (2nd ed.).* Upper Saddle River, NJ: Prentice Hall.

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. E. (1993). Three principals for models of category learning. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.), *Categorization by human and machines: The psychology of learning and motivation* (Vol. 29, pp. 57-90). San Diego, CA: Academic Press.

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.

Matsuka, T (2002). Attention processes in computational models of categorization. Unpublished Doctoral Dissertation. Columbia University, NY.

Matsuka, T. & Corter, J. E. (2003a). Neural network modeling of category learning using Generalized Radial Basis Functions. Paper presented at 36th Annual Meeting of the Society of Mathematical Psychology. Ogden, UT.

Matsuka, T. & Corter, J. E. (2003b). 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 (2004). Modeling category learning with stochastic optimization methods. In *Proceeding of International Conference on Cognitive Modelling.* Pittsburgh, PA

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

Minda, J.P. & Smith, J.D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 275-292.

Nosofsky, R.M. & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 924-940.

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

Poggio, T. & Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science, 247,* 978-982.

Rosseel, Y. (1996). Connectionist models of categorization: A statistical interpretation. *Psychologica Belgica, 36,* 93-112

Shanks, D.R. (1991). Categorization by a connectionist Network. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 433-443.

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