In Proceedings of the Sixth International Conference on Cognitive Modeling, 3370-371. Mahwah, NJ: Lawrence Earlbaum.

Irresistibly Attractive Fruitless Feature Dimensions

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

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

Stephen Jose Hanson (jose@psychology.rutgers.edu)
Psychology Department, Rutgers University - Newark

Although selective attention allocation has been suggested to be one of the most important processes implemented in the recent computational models of category learning (e.g., Kruschke, 1992), the models' predictions on attention allocation have been virtually ignored by cognitive modeling researchers. Rather almost all modeling studies had focused solely on the models' capabilities in reproducing observed learning curves or classification response patterns, and thus the models' descriptive validities on their selective attention allocation processes have remained still mostly untested. Though they did not directly evaluate the descriptive validities of attention processes in computational models of category learning, Matsuka, Corter, and Markman (2004) conducted a set of simulation studies indicating that information on predicted patterns of attention allocation can be informative in differentiating models. In that study, three recent models of category learning, including ALCOVE (Kruschke, 1992) and SUSATIN (Love & Medin, 1998, Love, Medin, Gureckis, 2004), were compared in multiple aspects, including predicted attention allocation patterns. Matsuka et al. (2004) found that the models perform comparably in reproducing the observed classification response profiles but gave markedly different predictions on selective attention allocation patterns. The results of the simulation studies also indicated the models have different tendencies in their attention allocation patterns. ALCOVE has a tendency to be attracted to a feature dimension that consists of unique elements which help differentiate each exemplar. SUSTAIN on the other hand, has a tendency to be attracted to a feature dimension whose elements are less diverse or more homogeneous.

If these tendencies are generalizable, then their descriptive validities may become questionable in some cases. For example, in order to categorize humans into either male or female, ALCOVE would pay good amount of attention to each person's finger print (i.e., feature dimension consists of unique elements), and SUSTAIN would pay good amount of attention to the fact the they are humans (i.e., feature dimension consist of a constant element). These tendencies, however, in less extreme cases, might be realistic tendencies in humans' "irrational" cognitive processes. In any case, we first should examine the models' mechanisms that are associated with such tendencies in order to evaluate the models' descriptive validities. Specifically, in the present study, we extend Matsuka et al. (2004) to investigate those tendencies in attention allocation associated with two prominent models of category learning, namely ALOCE (Kruschke, 1992) and SUSTAIN (Love et al, 1998, 2004) with a simulation study using an artificial stimulus set.

Table 1: Artificial stimulus set used in the present study.

	Category A				Category B			
Dim 1	1	1	1	1	0	0	0	0
Dim 2	1	1	1	1	1	1	1	1
Dim 3	1	2	3	4	5	6	7	8

Simulations

We investigated aforementioned two computational models of category learning in their predicted attention allocations in a simulated classification task. In particular, we created an artificial stimulus set consisting of two different types of ineffective feature dimensions that are suggested to attract the models' attention in different ways. Table 1 shows the stimulus set used in present study. Note that Dimension 1 is the most informative dimension, perfectly correlated with the category membership. Dimensions 2 is constant and thus non-diagnostic at all, but SUSTAIN is expected to pay a good amount of attention to this dimension. Dimension 3 on the other hand can be informative, but does not allow people to have abstract concepts about the categories (because all elements are unique). Thus, in real word categorization task, Dim3 is not expected to be attended or at least is expected to be less attended than Dim 1. However, ALCOVE is expected to pay a good amount of attention to Dim3, if the claim by Matsuka et al (2004) is valid. Note that all feature values in Dim3 are treated as nominal value differentiating from each element within the dimension, and thus their numeric differences do not have any meaning.

In order to investigate general tendencies in their attention allocation patterns for two different types of ineffective feature dimension, 10,000 simulated subjects with randomly assigned parameter values were trained to classify the artificial stimulus set. The ranges of parameters were [0.1 10] for c and ϕ, [0.001 1] for the two learning rates for ALCOVE. For SUSTAIN the ranges were set at [0.1 10] for β, d, and r, and [0.001 1] for the learning. Note that we tested the original version of SUSTAIN (Love & Medin, 1998). Readers are advised to refer to Kruschke (1992) and Love et al. (1998 & 2004) for the details of ALCOVE and SUSTAIN.

The two models were run in a simulated training procedure to learn the correct classification responses for the stimulus set. The models were run for 20 blocks of training, where each block consisted of a complete set of the training instances.

Results and Discussion

The predicted *relative* attention allocation by ALOCVE and SUSTAIN are summarized in Table 2. As expected ALCOVE, on average, paid the greatest amount of attention to Dimension 3 (i.e., one with all unique elements). In contrast, SUSTAIN allocated the same amount of attention to the diagnostic dimension (i.e., Dim1) and the constant dimension (Dim2). Surprisingly, although there were 10,000 randomly chosen parameter configurations, there was virtually no variability in SUSTAIN's attention distribution pattern (i.e., the standard deviations were less than 0.001 for all dimensions). In other words, regardless of its parameters, SUSTAIN always allocated its attention in the same manner. In contrast, ALCOVE shows greater variability, and ALCOVE did not always pay more attention to Dim 3 than Dim 1 (see Figure 1). Approximately 60% of simulated ALCOVE subjects paid more attention to Dim3.

But what causes ALCOVE to pay more attention to Dimension 3 than Dimension 1? The bivariate scatter plots (or correlation coefficients) suggest when the learning rate for association weights (LRW) is high, ALCOVE is more likely to pay greater amount attention to Dim 3 (Figure 1, top panels), and when it is low, ALCOVE is more likely to pay lesser amount of attention to the dimension. This is because when the LRW is high, ALCOVE learns to form a strong association between each exemplar to the correct category node very quickly, which in turn, during learning phase, sends a stronger signal to the exemplar node that has identical features to the current input stimulus. Then, the feature dimension that distinguishes the exemplar (with strong feedback signal) from other exemplars receives more amount of attention weight update, resulting in distributing the most attention to the unique dimension (Dim3)

This result shows that ALCOVE can model at least two types of learners; domain experts, who are very familiar with many exemplars in a given domain and learn to form associations between exemplars and categories quickly; and ordinary people, those who have more abstract concept on the domain and do not learn about the associations as quickly as the domain experts. Thus, this ALCOVE's tendency seems to be descriptively valid in some extent and/or cases. It is rather interesting, however, that the learning rate for attention is less sensitive to this ALCOVE's tendency in attention allocation than LRW.

The next issue to be discussed is why SUSTAIN pays equal amount of attention to Dim1 and Dim2. This can be attributed to the way SUSTAIN utilizes its reference points (i.e., prototypes and exceptions) in learning. SUSTAIN utilizes only the single most activated reference point (RP) for learning (and classification). In SUSTAIN, the update in attention strength for each dimension is determined as a function of the distance from the most activated RP and the current input stimulus in that dimension (i.e., smaller the dimensional distance, more attention to be allocated to that dimension). SUSTAIN's winner-take-all type of RPs-utilization (or more precisely, receiving no information on each dimension's category-diagnosticity from other RPs during learning) can be problematic, because as we have seen it does not effectively reduce attention distribution to a non-diagnostic constant dimension. This is because regardless of stimulus and/or category structure, if there is a constant dimension, then the constant dimension always has zero distance for all RPs in the corresponding dimension. Since only one most activated RP is used for learning, SUSTAIN, regardless of particular RPs to be most activated, always finds the smallest distance (i.e., zero) in the constant dimension and thus always misperceives the comparative advantage of the constant dimension (as compared with other dimension) resulting in paying a great amount of attention to the dimension.

The most apparent limitation of the present study is that we do not have empirical evidence about how people would utilize those dimensions. However, we, in some extent, showed the mechanisms of ALCOVE and SUSTAIN's tendencies in attention allocations. Empirical data on attention allocation and simulation studies with several different stimulus structures can be very fruitful for this line of work.

Table 2: Predicted relative attention allocations

	ALCOVE		SUSTAIN	
	Mean	Stddev	Mean	Stddev
D1	0.335	0.199	0.477	0.000
D2	0.190	0.125	0.477	0.000
D3	0.475	0.294	0.046	0.001

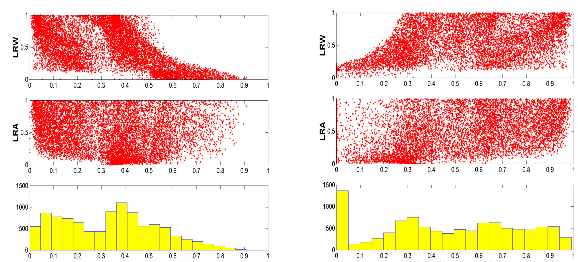

Figure 1: Predicted attention allocation to Dimensions 1 (left bottom panel) & 3 (right bottom panel), and their relationships to the learning rate (LR) for association weights (top panels) and LR for attention (middle panels).

References

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

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.

Love, B.C., Medin, D.L., & Gureckis, T.M. (2004). SUSTAIN: A network model of human category learning. *Psychological Review*.

Matsuka, T., Corter, J. E. & Markman, A. B. (2004). Allocation of attention in neural network models of categorization. Under review..