Modeling Item and Category Learning

Bradley C. Love (loveb@nwu.edu)
Department of Psychology; 2029 Sheridan Road
Evanston, IL 60208-2710 USA

Douglas L. Medin (medin@nwu.edu)
Department of Psychology; 2029 Sheridan Road
Evanston, IL 60208-2710 USA

Abstract

SUSTAIN (Supervised and Unsupervised STratified Adaptive Incremental Network) is a network model of human category learning. SUSTAIN is a three layer model where learning between the first two layers is unsupervised, while learning between the top two layers is supervised. SUSTAIN clusters inputs in an unsupervised fashion until it groups input patterns inappropriately (as signaled by the supervised portion of the network). When such an error occurs, SUSTAIN alters its architecture, recruiting a new unit that is tuned to correctly classify the exception. Units recruited to capture exceptions can evolve into prototypes/attractors/rules in their own right. SUSTAIN's adaptive architecture allows it to master simple classification problems quickly, while still retaining the capacity to learn difficult mappings. SUSTAIN also adjusts its sensitivity to input dimensions during the course of learning, paying more attention to dimensions relevant to the classification task. SUSTAIN successfully fits item and category learning data from Medin, Dewey, and Murphy (1983). SUSTAIN's performance on other data sets is discussed. SUSTAIN is compared with other models of category learning.

Introduction

Some categories have a very simple structure, while others can be complex. Accordingly, learning how to properly classify items as members of category "A" or "B" can be almost trivial (e.g., the value of a single input dimension determines membership) or can be so difficult that no regularity is discovered (e.g., rote memorization of every category member is required to determine membership).

Classifications are harder to master when the decision boundary (in a multi-dimensional space of possible inputs) is highly irregular and when there are multiple boundaries (e.g., all the members of category "A" do not fall inside one contiguous region of the input space). Very simple learning models will fail to master difficult categorizations with complex boundaries. For instance, a purely linear model, like the perception (Rosenblatt, 1958) will not be able to master a classification where the mapping is not linearly separable.

Interestingly, a complex nonlinear model, such as a backpropagation model (Rumelhart, Hinton, & Williams, 1986) with many hidden units, can learn complex decision boundaries but will perform poorly on a simple problem (e.g., a problem where the decision boundary is linear). In such cases, the more complex model will generalize poorly by over-fitting the training data. Thus, making a model too powerful or too weak is undesirable. Geman, Bienenstock, and Doursat (1992) termed this tradeoff between data fitting and generalization as the bias/variance dilemma. In brief, when a network is too simple it is overly biased and cannot learn the correct boundaries. Conversely, when a network is too powerful, it masters the training set, but the boundaries it learns are somewhat arbitrary and are highly influenced by the training sample, leading to poor generalization.

Unfortunately, many learning models require that the number of intermediate level units be specified in advance: the number of hidden units in backpropagation, the number of codebook vectors in LVQ (Learning Vector Quantization, Kohonen (1990)), and the number of radial basis functions in the fixed architecture version of Poggio and Girosi's (1990) regularization network and ALCOVE (attention learning covering map, Kruschke (1990)). The problem may not be avoidable by treating the number of intermediate units as an additional parameter. The environment a model is embedded in could change and alter the nature of the decision boundaries. Also, certain architectures may be preferable at certain stages of the learning process. For example, Elman (1994) provides computational evidence (which seems in accord with findings from developmental psychology) that beginning with a simple network and adding complexity as learning progresses improves overall performance.

Models with an adaptive architecture (like SUSTAIN), do not need to specify the number of intermediate units prior to learning. Still adaptive architecture models are not without problems. Some models grow in an unconstrained fashion, adding an intermediate unit every time an item is presented (e.g., the adaptive architecture version of Poggio and Girosi's (1990) model and ALCOVE). Clearly, these models (as specified) have prohibitive space requirements (with an increase in time requirements when run on a serial computer) and may be psychologically implausible. While these model have adaptive architectures, architectural changes do not occur in response to how the learning process is unfolding. Also, one could argue that the solutions these models derive lack elegance and are difficult to interpret.

Other methods do make architectural changes in response to how learning is progressing. Pruning methods begin with a large network and remove units as learning progresses (Karnin, 1990). In practice, this method prove inefficient (it begins with a large network) and the algorithm often terminates with a medium sized network when a simpler network would be better suited to the learning problem. Another problem is that the modeler must decide how large the network should be in advance.

Instead, other models (including SUSTAIN) begin with a small network and expand the network when necessary. Most

methods expand the network when overall error (the difference between desired and observed output) is high. For example, the cascade-correlation model (Fahlman & Lebiere, 1990) expands the network vertically with additional intermediate layers, creating higher-order feature detectors. Other models expand horizontally when error is high (Ash, 1989; Azimi-Sadjadi, Sheedvash, & Trujillo, 1993).

Unlike the aforementioned models, SUSTAIN does not accrue units based on overall error. Instead, SUSTAIN adds a new intermediate level unit when the unsupervised part of the network clusters input patterns in a manner deemed inappropriate by the supervised part of the network. This happens when two input patterns (that differ) belong to the same cluster and the differences between the two input patterns proves critical for successfully mastering the classification. When such an error occurs, SUSTAIN splits the cluster into two clusters by adding an intermediate unit. Adding units in SUSTAIN is psychologically motivated by the intuition that people ignore differences when they can (a bias towards simple solutions), but will note differences when forced to by environmental feedback.

Another aspect of networks that is usually fixed, but should vary depending on the nature of the learning problem, is the activation function of an intermediate level unit. In backpropagation networks, the steepness of a hidden unit's sigmoidal shaped activation function is set as a parameter. In models where the intermediate level units are viewed as receptive fields (e.g., Poggio & Girosi, 1990; Kruschke, 1992), the shape of a unit's receptive field is set as a parameter.

Intermediate level units in SUSTAIN have multiple receptive fields (one for each input dimension) and treat the shape of a receptive field as something that should be learned, rather than as a parameter. SUSTAIN assumes that receptive fields are initially broadly tuned and are adjusted during the course of learning to maximize the receptive field's response to inputs. Intermediate units with peaked (narrow) receptive fields can be described as highly focused. Receptive fields that develop tighter tunings are capable of stronger responses to stimuli (see Figure 1). As an outcome of learning how to perform a classification, SUSTAIN learns which dimensions of the stimuli are relevant and should be attended to. Conceiving of attention as enhancing the tuning of cells is consistent with current work on the neural basis of attention (Treue & Maunsell, 1996).

An Overview of SUSTAIN

SUSTAIN consists of three layers: input, subcategory, and category. Input layer units take on real values to encode information about the environment (e.g., the encoding of an item that needs to be classified as belonging to category "A" or "B"). Units in the subcategory layer (the intermediate layer) encode the prototypes and exceptions of the category units. Subcategory units compete with one another to respond to patterns at the input layer with the winner (the subcategory unit that is most active) being reinforced. Weights are adjusted according to the Kohonen unsupervised learning rule for developing self-organizing maps (Kohonen, 1984). When a subcategory unit "wins" the centers of its receptive fields (there is a receptive field for each input dimension) move in the direction of the input pattern, minimizing the distance between the centers and the input pattern. This method is similar to a number of clustering techniques used for classification and pattern recognition, such as maximum-distance, K-means, and isodata (Tou & Gonzalez, 1974; Duda & Hart, 1972).

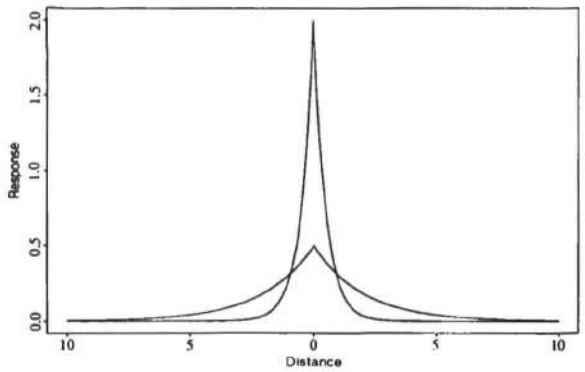

Figure 1. Both units respond maximally when a stimulus appears in the center of their receptive field (a .5 response for the broadly tuned unit; a 2.0 response for the tightly tuned unit). Compared to the broadly tuned unit, the tightly tuned unit's response is stronger to stimuli close to the center and is weaker for stimuli farther from the center (the crossover point occurs at a distance from center of .9 (approximately).

One novel aspect of our implementation is that this unsupervised learning procedure is combined with a supervised procedure. When a subcategory unit responds strongly to an input pattern (it is the winner) and has an excitatory connection to the inappropriate category unit (i.e., the subcategory unit predicts "A" and the correct answer is "B"), the network shuts off the subcategory unit and recruits a new subcategory unit that responds maximally to the misclassified input pattern (i.e., the new unit's receptive fields are centered upon the input pattern).[1] This process continues with the new unit competing with the other subcategory units to respond to input patterns with the position of the winner's receptive fields being updated, as well as its connection to the category units by the delta learning rule (Rumelhart et al., 1986). At a minimum, there must be as many subcategory units as category units when category responses are mutually exclusive.

Previous proposals that bear some resemblance to SUSTAIN include counterpropagation networks (Hecht-Nielsen, 1988) which are multilayer networks where the Kohonen learning rule is used for the bottom two layers. Simpson has explored a supervised version of the Kohonen network where the model does not determine which cluster is the winner, but is told (Simpson, 1989). This change greatly speeds up learning. Interestingly, our approach to clustering is not properly characterized as being either supervised or unsupervised. Clustering is unsupervised unless the network makes a serious clustering error. A serious error leads to the creation of a new cluster; otherwise learning at the subcategory layer is completely unsupervised.

Another interesting aspect of SUSTAIN's subcategory units is that in addition to adjusting the centers (i.e., the posi-

[1] Initially the network only has one subcategory unit that is centered upon the first input pattern.

tion) of their receptive fields, the sensitivities (i.e., the shape) of their receptive fields also are adjusted in response to input patterns. Input units (i.e., dimensions of the input pattern) that provide consistent evidence (i.e., the position of the subcategory units' receptive fields for that dimension does not have to be adjusted often), develop tighter tunings (see Figure 1). These more reliable input dimensions receive more attention. SUSTAIN uncovers (and explicitly represents) which dimensions are relevant for classification.

An Illustration of SUSTAIN's Operation

Consider categorizing people into two groups: those who ride motorcycles and those who don't. For the motorcyclists category, SUSTAIN might form an initial cluster that responds to input patterns representing young, adventurous men. When an input pattern representing a 40 year old recently divorced wealthy man that rides a motorcycle is presented to the network, it activates a subcategory unit associated with non-motorcyclists more strongly than it activates the subcategory unit that represents young male motorcycle riders. To remedy this situation, SUSTAIN creates a new unit that is centered upon the 40 year old divorced man. It turns out that this exception correctly classifies a number of other input patterns (the cluster can be labeled "mid-life crisis"). At this point, the motorcyclists category contains two distinct prototypes ("young male" and "mid-life crisis"). If the network was presented with a grandmother that likes to bungee jump, the network would probably predict she doesn't ride a motorcycle, but if it turns out she does, a new subcategory unit would be created to capture this exception. Other adventurous older women that are similar to the grandmother will also activate this unit, perhaps turning this exception into prototype.

During the learning process, tunings for each input dimension are developed. Dimension like "hair color" would be broadly tuned whereas dimensions like "has two legs" and "is a free spirit" may develop sharper tunings and be more influential in selecting the winning subcategory unit.

Mathematical Formulation

Receptive fields have an exponential shape with a receptive field's response decreasing exponentially as distance from its center increases:

$$\alpha(\mu) = \lambda e^{-\lambda\mu} \tag{1}$$

where λ is the tuning of the receptive field, and μ is the distance of the stimulus from the center of the field. Arguments for activation dropping off exponentially can be found in (Shepard, 1987).

While receptive fields with different λ have different shapes, for any λ, the area "underneath" a receptive field is constant:

$$\int_0^\infty \alpha(\mu)d\mu = \int_0^\infty \lambda e^{-\lambda\mu}d\mu = 1. \tag{2}$$

For a given μ, the λ that maximizes $\alpha(\mu)$ can be computed by differentiating:

$$\frac{\partial\alpha}{\partial\lambda} = e^{-\lambda\mu}(1-\lambda\mu). \tag{3}$$

These properties of exponentials prove useful in formulating SUSTAIN.

The activation of a subcategory unit is given by:

$$A_{H_j} = \frac{\sum_{i=1}^{n}(\lambda_i)^r e^{-\lambda_i\mu_{ij}}}{\sum_{i=1}^{n}(\lambda_i)^r} \tag{4}$$

where n is the number of input units, λ_i is the tuning of each subcategory unit's receptive field for the ith input dimension, μ_{ij} is the distance between the center of subcategory unit j's receptive field for the ith input unit and the output of the ith input unit (distance is simply the absolute value of the difference of these two terms),[2] and r is an attentional parameter (always nonnegative). When r is high, input units with tighter tunings (units that seem relevant) dominate the activation function. Equation 4 sums the responses of the receptive fields for each input dimension and normalizes the sum. The activation of a subcategory unit is bound between 0 (exclusive) and 1 (inclusive).

Subcategory units compete to respond to input patterns and in turn inhibit one another. When many subcategory units are strongly activated, the output of the winning unit is less. Units inhibit each other according to:

$$O_{H_j} = \frac{(A_{H_j})^\beta}{\sum_{i=1}^{m}(A_{H_i})^\beta} A_{H_j} \tag{5}$$

where β is the lateral inhibition parameter (always nonnegative) and m is the number of subcategory units. When β is small, competing units strongly inhibit the winner. When β is high the winner is weakly inhibited. Units other than the winner have their output set to zero.[3]

After feedback is provided by the "experimenter", if the winner predicts the wrong category, its output is set to zero and a new subcategory unit is recruited:

$$\text{for all } j \text{ and } k, \text{ if } (t_k O_{H_j} w_{jk} < 0), \text{ then recruit a new unit} \tag{6}$$

where t_k is the target value for category unit k and w_{jk} is the weight from subcategory unit j to category unit k. When a new unit is recruited its receptive fields are centered on the misclassified input pattern and the subcategory units' activations and outputs are recalculated.

If a new subcategory unit is not created, the centers of the winner's receptive fields are adjusted:

$$\Delta w_{ij} = \eta(O_{I_i} - w_{ij}) \tag{7}$$

where η is the learning rate, O_{I_i} is the output of input unit i. The centers of the winner's receptive fields move towards the input pattern according to the Kohonen learning rule. This

[2]Distance must be calculated in a different manner when two or more input dimensions are integral (e.g., lightness and saturation in color perception). In such cases, the Euclidean distance between the expected pattern and the observed pattern is calculated. Integral input dimensions also share a common λ. The reader can consult Shepard (1964) and Nosofsky (1987) for more information on the metric properties of integral dimensions.

[3]The model (as specified) can have multiple winners. For instance, there could always be two winners. More complex schemes could also be considered for determining the number of winners. We do not explore any of these possibilities because they are less conceptually clear and the data does not demand it.

learning rule centers the prototype (i.e., the cluster's center) amidst the members of the prototype.

Using our result from Equation 3, receptive field tunings are updated according to:

$$\Delta\lambda_i = \eta e^{-\lambda_i \mu_{ij}} \left(1 - \lambda_i \mu_{ij}\right). \tag{8}$$

Only the winning subcategory unit updates the value of λ_i. Equation 8 adjusts the shape of the receptive field for each input so that each input can maximize its influence on subcategory units. Initially, λ_i is set to be broadly tuned. For example, if input unit i takes on values between -1 and 1, the maximum distance between the ith input unit's output and the position of a subcategory unit's receptive field (for the ith dimension) is 2, so λ_i is set to .5 because that is the optimal setting of λ_i for μ equal to 2 (i.e., Equation 8 equals zero). Under this scheme, λ cannot become negative during training.

Activation is spread from the winning subcategory unit to the category units:

$$A_{C_k} = O_{H_j} w_{jk} \tag{9}$$

where A_{C_k} is the activation of the kth category unit and O_{H_j} is the output of the winning subcategory unit.

The output of a category unit is given by:

$$\begin{aligned} &\text{if } (C_k \text{ is nominal and } |A_{C_k}| > 1), \text{ then } O_{C_k} = \frac{A_{C_k}}{|A_{C_k}|} \\ &\text{else } O_{C_k} = A_{C_k} \end{aligned} \tag{10}$$

where O_{C_k} is the output of the kth category unit. If the feedback given to subjects concerning C_k is nominal (e.g., the item is in category "A" not "B"), then C_k is nominal. Kruschke (1992) refers to this kind of teaching signal as a "humble teacher" and explains when its use is appropriate.

When a subcategory unit is recruited, weights from the unit to the category units are set to zero. The one layer delta learning rule (Rumelhart et al., 1986) is used to adjust weights these weights:

$$\Delta w_{jk} = \eta(t_k - O_{C_k}) O_{H_j} \tag{11}$$

where t_k is the target value (i.e., the correct value) for category unit k. The target value is analogous to the feedback provided to human subjects. Note that only the winner will have its weights adjusted since it is the only subcategory unit with a nonzero output.

The following equation determines the response probabilities (for nominal classifications):

$$Pr(k) = \frac{(O_{C_k} + 1)^d}{\sum_{i=1}^{p} (O_{C_i} + 1)^d} \tag{12}$$

where d is a response parameter (always nonnegative) and p is the number of category units. The category unit with the largest output is almost always chosen when d is large. In Equation 12, one is added to each category unit's output to avoid performing calculations over negative numbers. The Luce choice rule is a special case ($d = 1$) of this decision rule (Luce, 1963).

Empirically Testing SUSTAIN

SUSTAIN has successfully fit Shepard et al.'s (1961) classic experiments on the time course of human category learning (Love & Medin, 1998).[4] In Shepard et al.'s study, subjects assigned a stimulus to either category "A" or "B" and feedback was provided. Six different assignments of objects to categories were tested with the six problems varying in difficulty (Type I was the easiest to master, Type VI the hardest). For example, the Type I problem can be solved by attending to only one input dimension (e.g., color), while Type VI requires attending to all three dimensions (color, shape, and size) and has no regularities across any pair of dimensions.

SUSTAIN successfully fit subjects' learning curves and its solution was readily interpretable. SUSTAIN recruited more subcategory units for the more difficult problems. For example, the most common solution for the Type I problem was to create one unit for each category. Type VI has no regularities that can be exploited, forcing SUSTAIN to "memorize" each stimulus (i.e., SUSTAIN devoted a subcategory unit to each input pattern).

The Type VI problem is in some ways equivalent to identification learning while the Type I problem seems like a "pure" categorization problem (there is a simple criteria for membership, the categories are very cohesive). The relative difficulty of the Type VI problem suggests (incorrectly) that identification learning is always more difficult than categorization learning, or more generally, that classification becomes easier at increased levels of abstraction. Contrary to this conclusion, there are striking instances where identification precedes categorization.

For example, Medin et al. (1983) found that people are faster to associate a unique names to photographs of nine female faces than they are to categorize the photographs into two categories. The logical structure of the two categories is shown in Table 1. One possible explanation for the relative ease of identification learning is that the stimuli used in Medin et. al (1983) were rich and distinct, varying along many dimensions not listed in Table 1, such as the shape of the face, the type of nose, etc.. This *idiosyncratic* information makes each stimulus item more distinct.

SUSTAIN correctly predicts that the relative rates of identification and categorization learning interact with the nature of the stimuli (with the same parameter values used to model Shepard et al. (1961): $\eta = .1$, $\beta = 1.0$, $r = 3.5$, and $d = 8.0$). Specifically, when the stimuli are highly distinct, identification learning is faster than categorization. The properties of SUSTAIN that give rise to this behavior will be discussed after simulation results are presented for Medin et al. (1983).

Modeling Medin et al. (1983)

Subjects were assigned to one of a number of learning conditions. Here, we focus on the First Name and Last Name condition. In the First Name condition subjects learned a separate label for each photograph, while in the Last Name condition only two labels were used. In both conditions, subjects trained until they correctly classified all nine items for two consecutive blocks or until they completed the sixteenth

[4] The data actually fit was from Nosofsky et al.'s (1994) replication.

Table 1
The logical structure of the two categories is shown. The four dimensions were hair color, smile type, hair length, and shirt color.).

Category A	Category B
1112	1122
1212	2112
1211	2221
1121	2222
2111	

Table 2
Human performance and SUSTAIN's (in parentheses).

Problem Type	Criterion	Overall
First Name	1.00 (1.00)	.84 (.73)
Last Name	.91 (.38)	.87 (.76)

learning block (a learning block consisted of presenting each item in Table 1 once in a random order). Feedback was provided.

The results from Medin et al. (1983) are shown in Table 2. Notice that every subject in the First Name condition reached criterion, while only 91% of subjects reached criterion in the Last Name condition. Also, accuracy overall was roughly equal, even though chance guessing favored the Last Name condition (i.e., pure guessing would result in 1/2 correct compared to 1/9 correct). When the First Name condition is rescored to account for guessing by scoring any label within the same category ("A" or "B") as correct, overall accuracy rises to 91%.

To fit SUSTAIN to the data, certain assumptions had to be made about the nature of the input representation. Because subjects were sensitive to the idiosyncratic information in each photograph, twenty additional input dimensions were added. Each of the twenty idiosyncratic dimensions consisted of nine input units with eight set to negative one and one set to positive one (e.g., for the dimension representing the type of nose, the fourth input unit of the nose dimension positive and the rest negative indicated that the face had the fourth type of nose, which is distinct from the other eight nose types). Put differently, each input dimension consisted of an attribute that could take on one of nine possible values. Each stimulus item had a unique value on each idiosyncratic input dimension. The input representation of the four dimensions listed in Table 1 also had nine input units per dimension, but only the first two units of these dimensions were ever positive.

To capture that the nine input units forming an input dimension are one functional unit, the λ values associated with each input unit forming an input dimension were averaged after each update. Also, μ (the distance from input) for each subcategory unit's receptive field was the sum across all input units belonging to the same dimension divided by 2 (only one of the nine units takes on a positive value, thus the maximum μ is 2, generalizing the binary feature case). To summarize, input units forming a dimension have a common μ and λ.

Table 3
Human performance and SUSTAIN's performance (in parentheses) with $d = 15$.

Problem Type	Criterion	Overall
First Name	1.00 (1.00)	.84 (.84)
Last Name	.91 (.89)	.87 (.86)

Table 4
SUSTAIN's final architecture and mean λ (2nd block).

Problem Type	Mean Subcategory Units	Mean λ
First Name	9.0	2.3
Last Name	6.8	1.5

SUSTAIN was run on each condition 10,000 times. SUSTAIN captured the major patterns in the data (see Table 2). SUSTAIN's quantitative fit of the data can be increased by setting the decision parameter d to 15 (see Table 3). The decision parameter determines the extent to which SUSTAIN stresses accuracy and can be viewed as outside the model.[5] Like people, SUSTAIN found it more natural to identify each stimulus than it did to associate several stimuli to a common label.

Table 4 shows the number of subcategory units recruited by SUSTAIN by condition. Notice that SUSTAIN recruited more units in the First Name condition than in the Last Name condition. SUSTAIN's tunings[6] are sharpest in the First Name condition, indicating that more input dimensions are relevant for classification in this condition (as would be expected). Interestingly, when SUSTAIN's input representation does not include idiosyncratic information, the Last Name condition (criterion: .50, overall: .77) is easier to master than the First Name condition (criterion: .00, overall: .20).

Why SUSTAIN favors identification over categorization in Medin et al. (1983)

Various factors conspire to cause SUSTAIN's performance to interact with the nature of the stimuli. Performance tends to improve when fewer subcategory units are recruited because units that respond to multiple stimulus items develop stronger associations with the category units. At odds with this preference for fewer units is a preference for highly specialized subcategory units. A subcategory unit specialized for a particular stimulus will respond very strongly to that stimulus. Another factor in favor of fewer subcategory units is that units inhibit each other. Making stimuli more distinctive alters the balance of these forces (leading to the observed interaction). The benefit of having fewer subcategory units diminishes with distinctive stimuli because distinctive inputs tend not to cluster as well together and subcategory units tend to inhibit each other less.

[5] All further discussion will focus on results from simulations with $d = 8$.

[6] SUSTAIN's mean tunings are reported after two learning blocks because some runs reached criterion at that point. Differences in tunings between the six conditions are magnified when later blocks are examined.

Discussion

SUSTAIN's ability to model both Shepard et al.'s (1961) and Medin et al.'s (1983) data highlight SUSTAIN's promise as a model of human category learning.[7] These results suggest that SUSTAIN may prove successful in explaining why certain categories are more natural or basic than others (Rosch et al., 1976). For example, if asked how one gets to work in the morning, one says, "I drive my *car*," as opposed to "I drive my *Buick*," or "I drive my *vehicle*." SUSTAIN offers an explanation for why a level of categorization is preferred. In the above example, the intermediary category *car* balances the need to create subcategory units that have a high degree of within cluster similarity and low degree of between cluster similarity while minimizing the total number of clusters (i.e., subcategory units). Also, SUSTAIN's shift towards lower level categories in the presence of more distinctive inputs may be in accord with shifts in preferred category level with expertise (Tanaka & Taylor, 1991).

Acknowledgments

B. C. Love was supported by the Office of Naval Research under the National Defense Science and Engineering Graduate Fellowship Program.

References

Ash, T. (1989). Dynamic node creation in backpropagation networks. *Connection Science*, *1*(4), 365–375.

Azimi-Sadjadi, M. R., Sheedvash, S., & Trujillo, F. O. (1993). Recursive dynamic node creation in multilayer neural networks. *IEEE Transactions on Neural Networks*, *4*(2), 242–256.

Billman, D. & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*(2), 458–475.

Carpenter, G. A. & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Proc.*, *37*, 54–115.

Duda, R. O. & Hart, P. E. (1972). *Pattern Classification and Scene Analysis*. New York: Wiley.

Elman, J. L. (1994). Implicit learning in neural networks: The importance of starting small. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing.*, pp. 861–888. Cambridge, MA: MIT Press.

Fahlman, S. E. & Lebiere, C. (1990). The cascade-correlation learning architecture. In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems 2. Proceedings of the 1989 Conference*, pp. 524–532. San Mateo, CA: Morgan Kaufmann.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, *4*(1), 1–58.

Hecht-Nielsen, R. (1988). Applications of counterpropagation networks. *Neural Networks*, *1*(2), 131–139.

[7]SUSTAIN has also been successfully applied to Billman & Knutson's (1996) unsupervised learning data (Love & Medin, 1998). One additional parameter is added to SUSTAIN that governs when a new subcategory unit is recruited. When the winning subcategory unit's output is below some threshold, a new subcategory unit is added. The ART (Adaptive Resonance Theory) model (Carpenter & Grossberg, 1987) of unsupervised category learning operates in a similar fashion.

Karnin, E. D. (1990). A simple procedure for pruning backpropagation trained neural networks. *IEEE Transactions on Neural Networks*, *1*(2), 239–242.

Kohonen, T. (1984). *Self-Organization and Associative Memory*. Berlin, Heidelberg: Springer. 3rd ed. 1989.

Kohonen, T. (1990). Improved versions of Learning Vector Quantization. In *Proc. IJCNN-90-San Diego, Int. Joint Conf. on Neural Networks*, Vol. I, pp. 545–550 Piscataway, NJ. IEEE Service Center.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44.

Love, B. C. & Medin, D. L. (1998). SUSTAIN: A model of human category learning. Submitted.

Luce, R. D. (1963). Detection and recognion. In R. D. Luce, R. R. Busg, & E. Galanter (Eds.), *Handbook of Mathematical Psychology*, pp. 103–189. New York: Wiley.

Medin, D. L., Gerald, G. I., & Murphy, T. D. (1983). Relationships between item and category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *9*, 607–625.

Nosofsky, R. M. (1987). Attention and learning processes in the indentification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*, 87–108.

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 & Cognition*, *22*, 352–369.

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

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage in the brain. *Psychological Review*, *65*, 386–408.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533–536.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54–87.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*(13, Whole No. 517).

Simpson, P. K. (1989). *Artificial Neural Systems*. Elmsford, NY: Pergamon Press.

Tanaka, J. W. & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder. *Cognitive Psychology*, *23*(2), 457–482.

Tou, J. T. & Gonzalez, R. C. (1974). *Pattern Recognition Principles*. Reading: Addison-Wesley.

Treue, S. & Maunsell, J. H. R. (1996). Attentional modulation of visual motion processing in cortical areas mt and mst. *Nature*, *382*(6591), 539–541.