Title

A Connectionist Model for Classification Learning - The lAK Model

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 17(0)

Author
Heydemann, Martin

Publication Date
1995

Peer reviewed

A Connectionist Model for Classification Learning - The IAK Model

Martin Heydemann
Institut fuer Psychologie
Technische Hochschule Darmstadt
Steubenplatz 12
D 64293 Darmstadt
Germany
heydemann@hrz1.hrz.th-darmstadt.de

Abstract

The connectionist model IAK (Information evaluation using configurations) for classification learning is presented here. The model can be placed between feature based (e.g. Gluck & Bower, 1988) and exemplar based models (e.g. ALCOVE, Kruschke, 1992). Specific to this model is that during learning, sets of input features are probabilistically sampled. These sets are represented, in a hidden layer, by configuration nodes. These configuration nodes are connected to output nodes that represent category labels. A further characteristic of the IAK model is a mechanism which enhances retrieval of information. Simulations with the IAK model can explain different phenomena of classification learning which have been found in experimental studies: A Type 2 advantage without dimensional attention learning observed by Shepard et al. (1961); a generalisation of prototypes; a generalization based on similarity to learned exemplars; a differential forgetting of prototypes and exemplars; a moderate interference (fan effect) caused by stimulus similarity; and the missing of catastrophic interference even in A-B/A-B_r-designs.

In classification tasks, stimuli are given that belong to different category names. Subjects have to classify old stimuli that have been presented during a prior learning phase and new stimuli. The classification depends on the involved stimuli and the degree of practice.

There are two different ways for connectionist modeling of classification learning. In feature-based models associations between single features of the stimuli and features of category names are formed during learning (e.g. Gluck & Bower, 1988; Estes et al. 1989). On the other side, exemplar-based models assume associations between representations of the whole stimulus and the category label, e.g. ALCOVE (Kruschke, 1992) or the context model (Medin & Schaffer, 1978). These models explain a lot of empirical phenomena of category learning.

The IAK-Model (IAK: Information evaluation using configurations) lies between feature-based and exemplar-based models. The IAK-Model exhibits two main properties:

- Associations between small sets of stimulus features (configurations) and category labels are learned.
- For optimizing recall, association weights between configurations and category labels are computed taking all currently competing weights into account.

Configurations of features are also used in the configural-cue model of Gluck and Bower (1988). In contrast to this model, IAK makes use of a probabilistic sampling process to select a small subset of configurations, thus avoiding a combinatory explosion of the number of configurations.

The Model

The IAK-Model requires three layers: Input nodes, configuration nodes and output nodes. Input nodes represent the features of the stimulus. Their activation is either 1 (on) or 0 (off). During learning, input nodes are connected to configuration nodes. A configuration node gets an activation of 1 if all input nodes that are connected to the configuration node are on. Input nodes and configuration nodes exhibit an all-or-none-activation characteristic. Output nodes represent categories for the classification tasks. Their activation values lie between 0 and 1.

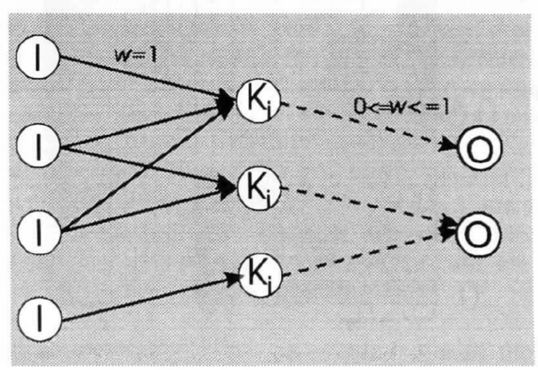

Figure 1. Connections in the IAK model. The connection from input to configuration node is either existent or absent. The connection from configuration to output node has a weight between 0 and 1.

Retrieval

An input pattern is activated and the system has to select a category represented by an output node by means of activation propagation from input nodes to output nodes. In classification tasks, the probability for selection of category *m* is:

$$p(m) = \frac{a_m}{\sum_{i=1}^{r} a_i} \qquad (1)$$

a_i is the activation of the i-th output node corresponding to the i-th category. r is the number of output nodes.

The output activations are computed as follows: First, configuration nodes are switched on if every connected input node is on. A single inactive connected input node causes the configuration node to remain inactive.

Second, the activation value a is computed for each output node:

$$a_j = \sum_{i=1}^{n} w_{ji}^{\tau} \cdot \left(1-(1-\delta)^{s_i}\right) \cdot (1-\delta)^{Ssum_{ji}} \qquad (2)$$

w_{ji} is the connection weight between the i-th configuration node and the j-th output node. s_i is the strength of the configuration node i. Strength values increase during learning. δ and ι are parameters of the IAK model. $Ssum_{ji}$ is computed by:

$$Ssum_{ji} = \sum_{k \in K_{ji}} s_k \qquad (3)$$

K_{ji} is the set of all configuration nodes connected to the output node j that have connection weights to j greater than w_{ji} ($w_{jk} > w_{ji}$ for all $k \in K_{ji}$).[1] The value of a lies between 0 and 1 because the strength values s are positive integers and the parameter values are limited to $0<\delta\leq 1$ and $\tau>0$.

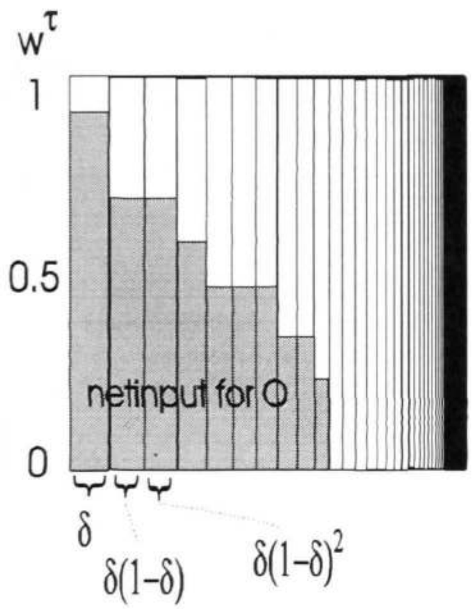

Figure 2. Illustration showing output node activation computing. The gray shaded area represents the activation value a. The maximum is 1. Each configuration node gets s columns; s is the strength of the configuration node. The configuration node with the maximum weight is placed in the first column, followed by the configuration node with the second biggest weight and so on.

These complicated equations prevent many relatively small values of w from obscuring larger values. The computation of a is illustrated in Figure 2.

Learning

Learning requires two computational steps:

1. Sampling of input nodes and strengthening corresponding configuration nodes.
2. Adjustment of weights from configuration nodes to output nodes.

For the first step, subsets of the active input nodes are sampled with a probabilistic procedure. Two parameters α and β control this process. α is the mean number of sets that are sampled in one learning trial and β influences the mean number of input nodes g in these sets. β is the linear slope for the probability gradient. The computation of β is illustrated in the following example. If there are 5 active input nodes I_1 to I_5 and $\alpha = 1.5$ and $\beta = -0.2$ then at least one subset is sampled and there is a probability of .5 that a second subset is sampled. The probability that the subset consists of 1 element is $p(g=1)=0.533$; $p(g=2) = 0.333$; $p(g=3) = 0.133$; $p(g=4) = 0$ and $p(g=5) = 0$. For instance, only one subset might be sampled with $g = 2$ and it might consist of I_2 and I_4. Now a configuration node is searched that has the connections to input nodes like the sampled subset. If a configuration node exists, its strength is incremented by 1. If not, then an unused configuration node is chosen with strength $s = 1$ and connections to the input nodes of the subset. If an unused configuration node does not exist then the forgetting process takes place to provide a node (see section Forgetting of configuration nodes).

In the second step, weights between all active configuration nodes and output nodes are adjusted. The weight w_{ji} gives the portion in which the output node j was a target in cases where the configuration node i was active. For instance, if configuration node i was active at 20 learning trials and at 15 of these 20 trials output node j was a target node then $w_{ji} = 0.75$.

Forgetting of Configuration Nodes

Forgetting of node connections is required in cases where the set of unused configuration nodes is exhausted and new ones are needed for learning. The following procedure is used repeatedly: A configuration node is randomly selected, and its strength is decremented by 1. Nodes with a strength of 0 are unassigned.

Parameters and Extended Versions of the IAK Model

This paper presents a reduced version of the IAK model. In the complete version during learning not only activating but also inhibiting connections from configuration nodes to output nodes are learned. These inhibitive links are rather selective and enhance the systems behavior in difficult

[1] If two or more active configuration nodes have the same weight to an output node then these nodes are treated as a single node with a summed strength value.

discrimination tasks. Another extension deals with configuration nodes for output nodes. These configuration nodes are suitable for learning complex response patterns and are not used in classification tasks. In the reduced model reported here, only four parameters are used α, β, δ, and τ, although others may be useful. For instance, a parameter is needed in Equation 1 to increase the activation values so that a medium value is not obsured by small ones. However, in the following simulations only qualitative results are reported and the parameters are kept at a minimum for better clarity of the model's mechanisms.

Applications

The Experiment of Shepard, Hovland, & Jenkins (1961)

The task. The stimuli vary on three binary dimensions: size (large vs. small), shape (square vs. triangle), and color (filled vs. empty). Four of them are assigned to category A the other to category B. There are six structurally different types of category assignment (see Figure 3).

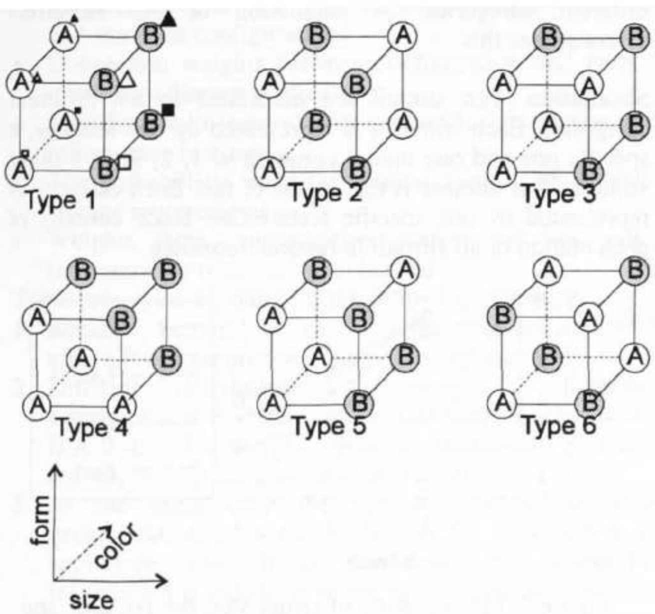

Figure 3. One example for the six types of stimuli assignments to categories.

Shepard et al. (1961; replicated by Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994). found the following ordering of difficulties in learning: Type 1 < Type 2 < (Type 3 to Type 5) < Type 6. The advantage of Type 2 compared to Types 3 to 5 is difficult to explain with connectionist models, unless the model has an explicit method to fade out irrelevant dimensions, e.g. ALCOVE (Kruschke, 1992) or DALR (Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994).

Method of simulation. The net consists of 2 output nodes representing categories A and B and 6 input nodes. Within one block each stimulus is presented twice in random order. Figure 4 shows the percentage of errors within each block from n = 400 simulations. The simulation replicates the ordering of type difficulties.

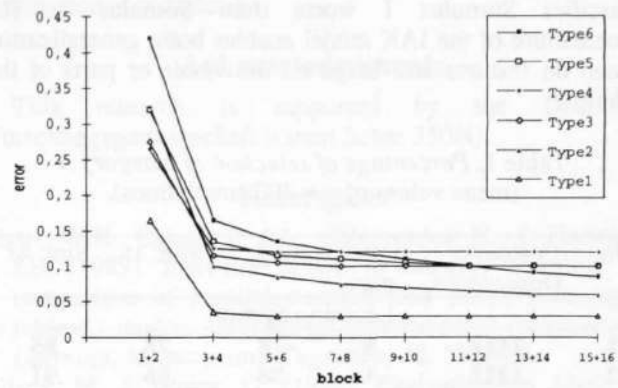

Figure 4. Mean error for 16 blocks of learning (Parameters: $\alpha = 5$; $\beta = 0$; $\delta = 0.007$; $\iota = 5$).

Results. In accordance to previous empirical data (Shepard et al, 1961; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994) there is no difference between linearly separable tasks (Type 4) to linearly non separable ones (Type 5). Also, an explicit mechanism of selective attention learning is not required by the IAK model to predict the advantage of Type 2 tasks compared to Type 3 to 5 tasks.

The Experiment of Medin and Schaffer (1978)

This experiment raises the question whether learning and generalization is based primarily on single features or alternatively on whole exemplars. If generalization is based on the sum of single feature-to-category associations, then the best classification should be found with the prototype stimulus for a category which consists of the features that are most typical for a category. An alternative assumption is that generalization may be based on similarity to whole exemplars.

In the experiment (exp. 2 of Medin and Schaffer, 1978) subjects had to classify stimuli with four binary dimensions. Stimuli 1 to 5 are learned as category A, Stimuli 6 to 9 are learned as B, and the remaining seven stimuli are only tested.

Method of simulation. For the simulation with the IAK model a net with eight input nodes and two output nodes is used. For each block of learning, the stimuli 1 to 9 are presented in random order. Table 1 compares the experimental results of the experiment of Medin and Schaffer (1978) with the results of the simulations after one (Sim: 1x) and four (Sim: 4 x) blocks of learning (Parameters: α=2; β=0; δ=0.01; ι=3).

Results. Three values should be compared in detail. Stimulus 12 is never learned but it is the prototype of

Category A and is classified best in the experiment and in the simulations. Stimulus 1 is more similar to the prototype of A than Stimulus 2, but Stimulus 1 is classified worse than Stimulus 2 in the experiment, because Stimulus 1 is highly similar to two stimuli (6 and 7) of the opposite Category B. The simulation with the IAK model also classifies Stimulus 1 worse than Stimulus 2. The architecture of the IAK model enables both: generalization based on features and based on the whole or parts of the whole.

Table 1. *Percentage of selection of category A* (mean values of n = 400 simulations).

No.	Values of Dimensions	Cate-gory	Exp.	Sim: 1x	Sim: 4x
		Learning Stimuli			
1	**1112**	**A**	**.78**	**.78**	**.85**
2	**1212**	**A**	**.88**	**.86**	**.91**
3	1211	A	.81	.91	.96
4	1121	A	.88	.79	.83
5	2111	A	.81	.79	.84
6	*1122*	B	.16	.38	.34
7	*2112*	B	.16	.37	.34
8	2221	B	.12	.22	.15
9	2222	B	.03	.12	.06
		Transfer Stimuli			
10	1221		.59	.60	.65
11	1222		.31	.46	.44
12	**1111**		**.94**	**.90**	**.96**
13	2212		.34	.46	.46
14	2121		.50	.53	.49
15	2211		.62	.62	.66
16	2122		.16	.25	.14

The Interaction in Forgetting Rates for Exemplars and Prototypes

One aspect of classification learning is the differential forgetting rate for exemplars and prototypes. In experiments (e.g. Homa et al. 1973) subjects learned to classify dot patterns that are randomly distorted versions of a prototype. They were tested with old exemplars from the learning phase, new unlearned exemplars and prototypes that were not shown during the learning phase. The main result was that forgetting is faster for exemplars than for prototypes.

Method of simulation. Per block, 18 exemplars in random order, 9 of Category A, 6 of B and 3 of C are presented. Each exemplar consists of eight features: four are specific for the exemplar, two are randomly selected from the category prototype and the last two are randomly selected from each of the competing categories. After three blocks of learning, old and new exemplars and prototypes are tested (Test 1). Forgetting is caused by a reduction of the configuration node strength (see section on forgetting of configuration nodes, page 4). After a forgetting rate of 80 percent the stimuli are retested (Test 2).

Table 2. Mean portion of errors (n=400 simulations; α=1; β=-0.4; δ=0.01; ι=5).

Test-Stimuli	Old	Prototype	New
	9 Learned Exemplars Category A		
Test 1	.06	.03	.33
Test 2	.23	.12	.41
	6 Learned Exemplars Category B		
Test 1	.19	.25	.73
Test 2	.45	.36	.66
	3 Learned Exemplars Category C		
Test 1	.24	.78	.93
Test 2	.65	.76	.86

Especially in Category B, the interaction between forgetting rates of exemplars and prototypes is evident. As in the experimental results: The forgetting for prototypes is slower than for old exemplars.

Interference-Effects

In the IAK model, as in other connectionist models, interference is caused by common features in stimuli from different categories. A simulation of the fan-effect demonstrates this.

Simulation. Ten stimuli are associated to ten different categories. Each stimulus is represented by two features, a specific one and one that is common to 1, 2, 3, or 4 other stimuli. This number is the degree of fan. Each category is represented by one specific feature.One block consists of presentation of all stimuli in random sequence.

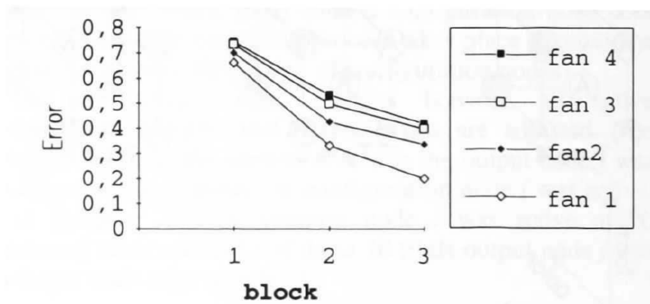

Figure 5. Mean portion of errors after the 1st, 2nd, and 3rd block of learning. (n=400 simulations; α=0.5; β=-0.3; δ=0.0001; ι=3).

Similar to empirical results, the main difference lies between degrees of fan from 1 to 2.

Size of interference. The degree of interference in the IAK model is similar to interference found with people. There is no catastrophic interference, in contrast to other connectionist models (McCloskey and Cohen, 1989). The following simulation of an A-B/A-B$_r$ design demonstrates this. First, List 1 with ten stimulus response associations (*a*

$k \rightarrow r_1$; $b\ k \rightarrow r_2$; $c\ k \rightarrow r_3$; $d\ k \rightarrow r_4$ ...)[2] is learned. *k* denotes the feature for the first context. Second, permutated combinations with all associations altered (List 2) are learned in a new context *m* (e.g.: $a\ m \rightarrow r_4$; $b\ m \rightarrow r_1$; $c\ m \rightarrow r_9$ ...). Test 1 is made after learning three blocks of List 1. Then follows learning of List 2 for three blocks and Test 2 is conducted. The simulation results in Table 3 show proactive and retroactive interference but no "catastrophic" interference.

Table 3: Mean portion of correct responses (n=400; α=5; β=-0.2; δ=0.01; ι=3).

	Teststimuli of List 1	Teststimuli of List 2
Test 1	.94	.00
Test 2	.75	.84

Discussion

This paper presents a reduced version of the IAK model. The model's powerful learning mechanism is nevertheless evident and applicable to more than classification tasks. This version of the IAK model demonstrates the following main learning mechanisms:

- Input features are sampled in an all-or-none manner and stored as configurations.
- Connection weights between configuration and output nodes are adjusted gradually.
- Specific (multi-feature) and unspecific (single-feature) information is stored.
- The probabilistic sampling process avoids unfulfillable storage requirements .
- Weights from configuration nodes that are valid indicators for retrieval are enhanced.

These principles are basis for the following properties:

1. Realistic, human-like results of learning are achieved after a few presentations of the learning material.
2. Difficult discrimination learning is possible. Interference is moderate but not catastrophic, even in A-B/A-B_r-transfer designs. Specific configuration nodes are responsible for good discrimination.
3. At the same time the system exhibits favorable generalization properties. If specific information is applicable, then it is used and the unspecific information is faded-out to prevent specific information from blurring. But, if there is no specific information, then unspecific information is increased in value providing a good generalization.

There are some structural similarities between the IAK model and RULEX (Nosofsky, Palmeri, & McKinley, 1994). Both models learn in a probabilistic way. Rules in RULEX may be compared to the binding of configurations to categories in IAK. In RULEX, it is easier to form simple rules with one feature than rules with two or more features (complicated rules and exceptions). This is the same in the IAK model, especially if the parameter β is negative. But there is one main difference between the models: In the IAK model a connection weight from a configuration to a category is kept even if inconsistent examples are encountered. However, inconsistent examples reduce the connection weight considerable. Thus, in IAK a configuration is only partially discarded. In RULEX, rules are discarded completely.

Acknowledgements

This research is supported by the Deutsche Forschungsgemeinschaft (Grant Schm 350/4).

References

Estes, W.K., Campbell, J.A., Hatsopoulos, N., & Hurwitz, J.B. (1989). Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 556-576.

Gluck, M. & Bower, G. (1988). Evaluation an adaptive network model of human learning. *Journal of Memory and Language, 27,* 166-195.

Homa, D., Cross, J., Cornell, D., Goldman, D., & Shwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology, 101,* 116-122.

Kruschke, J. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 97,* 225-252.

McCloskey, M. & N.J. Cohen (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G.H. Bower (Ed.) *The Psychology of Learning and Motivation (Vol. 24, pp. 109-165).* New York: Academic Press.

Medin, D.L. & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207-238.

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.

Nosofsky, R.M., Palmeri, T.J., & McKinley, S.C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53-79.

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

[2] a k $\rightarrow r_1$ denotes two input nodes a and k that are associated to the category node r_1.