An Unsupervised, Dual-Network Connectionist Model of Rule Emergence in Category Learning

Rosemary A. Cowell and Robert M. French
LEAD-CNRS, UMR5022, Université de Bourgogne, 21000 Dijon, France
{Rosemary.Cowell, Robert.French}@u-bourgogne.fr

Abstract

We develop an unsupervised "dual-network" connectionist model of category learning in which rules gradually emerge from a standard Kohonen network. The architecture is based on the interaction of a statistical-learning (Kohonen) network and a competitive-learning rule network. The rules that emerge in the rule network are weightings of individual features according to their importance for categorisation. Once the combined system has learned a particular rule, it de-emphasizes those features that are not sufficient for categorisation, thus allowing correct classification of novel, but atypical, stimuli, for which a standard Kohonen network fails. We explain the principles and architectural details of the model and show how it works correctly for stimuli that are misclassified by a standard Kohonen network.

Introduction

The categorisation of objects on the basis of their visual attributes is a cognitive capacity fundamental to our survival. The mechanisms underlying categorisation behavior in humans have been the subject of much theoretical and empirical work, both in adults and infants. Human adults, as well as infants above the age of around a year, are able to categorise objects based not only on the statistical structure of categories of observed objects, but also by making use of rules derived from that structure. Rules have the intrinsic advantage of radically reducing cognitive load: if an object can be categorised by paying attention to only a few of its features, instead of a great many, cognitive resources can be freed up for other tasks.

The ontological status of rules in a connectionist modeling framework has from the outset been a hotly debated topic (Seidenberg & McClelland, 1989; Pinker & Prince, 1988; Chalmers, 1990; Marcus et al., 1999; etc.). In this paper we have chosen a conciliatory point of view — namely, that rules do, indeed, have a distinct ontological status compared to purely statistical-learning mechanisms, but these rules, in general, must *emerge* from the "statistical" learning substrate.

A number of current models of category learning incorporate both a module for statistical learning of category structure and a rule module. The former gradually learns the statistical distributions of the perceptual attributes of objects in the world and uses this knowledge to determine the category membership of newly encountered objects. The rule module, on the other hand, has built-in rules capable of categorising these same objects directly. These models currently include, notably, ATRIUM (Erickson & Kruschke, 1998) and COVIS (Ashby et al., 1998). This distinction between statistical learning and rule-based learning parallels the distinction between exemplar models (Nosofsky, 1988; Kruschke, 1992; etc.) and prototype models (Rosch, 1978; Posner, 1986; etc.) of categorisation, as well as the distinction between implicit and explicit (i.e., verbal) categorisation strategies (Reber, 1967; Ashby et al., 1998, etc.).

It seems reasonable to assume that the acquisition of the rules underlying category structure should be possible through experience with stimuli from those categories. In other words, *it should be possible to extract knowledge of the rule automatically from knowledge about the statistical distribution of the perceptual characteristics of items in each category*. Current connectionist models of category learning that incorporate rule modules typically assume the *a priori* existence of these rules and model their application to the problem of object categorisation. These models do not, however, synthesize the rules themselves. For example, in ATRIUM (Erickson & Kruschke, 1998), the rule module contains an "off-the-shelf" rule for category membership; the stimulus dimension on which the rule is based is hard-wired, and the network must learn which values along that dimension are associated with each category. Similarly, in COVIS (Ashby et al., 1998), several pre-existing rules are hard-wired into the model's rule component and learning of the rule consists of selecting between available rules to find the one most appropriate to the current category structure.

Overview of the model

In what follows we will present a connectionist model of unsupervised category learning. This model consists of two interacting networks: a "statistical" network that learns the distributions of perceptual properties of the stimuli in each category and a "rule" network that derives its rules by continually monitoring the statistical network.

The statistical part of the network is a Kohonen network (Kohonen, 1982, 1993) and the rules emerge from a competitive network that monitors the Kohonen network. The Kohonen network self-organizes the inputs into a map in which representations of stimuli from the same category are clustered together. The competitive network monitors the Kohonen network as category learning proceeds and determines which input features are the most important in — in fact, *sufficient for* — determining category membership. This determination of a feature, or set of features, that is sufficient to determine category membership is what we mean by rule extraction.

We have chosen to implement our Kohonen network in a neurobiologically plausible manner, using leaky integrators, similar to an implementation described by Kohonen (1993). We suggest that processing of this type could occur in visual cortex and that a plausible candidate for the site of the competitive-learning algorithm used to model rule extraction could be pre-frontal cortex.

We will present a simulation that demonstrates the operation of the model. In particular, we will provide an example of an instance in which the statistical-learning component of the model (i.e., the Kohonen network) alone fails to generalise correctly from the learned category structure to a novel, atypical stimulus, whereas the combination of the statistical and rule-learning components of the model (i.e., the Kohonen and competitive learning networks, respectively) succeeds in correctly categorising the same stimulus.

Extraction of a rule

In the everyday categorisation of most commonly encountered classes of objects, the classifier can exploit the fact that the items belonging to a given category are likely to share a number of visual attributes: birds possess feathers, wings and a beak; tables almost always have legs and a flat surface; trees have a trunk, as well as leaves (or needles) during summertime.

A "rule" for category membership has traditionally been defined, in formal logic, as a necessary and sufficient condition — in this case, the presence of certain features in a particular combination — that unequivocally determines category membership. However, it has been recognised at least as far back as Wittgenstein (1951) that very few, if any, real-world categories have membership rules that meet this lofty standard. Therefore we can, in practice, use a "quasi-sufficient" condition for category membership as a "rule" for determining whether a given object is or is not a member of a real-world category. This simply means that, in general, the presence or absence of a particular feature (or set of features) is sufficient for determining category membership.

Rules of this nature might include: animals with feathers or beaks are birds; animals with gills are fish; land animals that weigh more than 5 tons are elephants; animals with opposable thumbs are primates; and so on. And while it is true that opossums, koalas and giant pandas also have opposable thumbs, and that the rule: "If X has a beak, X is a bird" caused early 19th century zoologists to think that duck-billed platypus specimens were a hoax, these rules are generally reliable and, most importantly, can be extracted from the feature statistics of primates and birds. This is precisely what our model does: *it identifies, for each category, the feature(s) whose presence is diagnostic of membership in that category*.

Further, it may well be that no single feature is sufficient for determining category membership, but a unique combination of features, each of which may be shared with other categories, will be sufficient to ensure correct category identification. For example, elephants live on land, as do lots of other animals, and weigh more than five tons, a property possessed by many species of whales. However, the combination of living on land AND weighing more than five tons is sufficient for correct category identification. Our model is also capable of extracting this type of conjunctive combination of features for category identification.

We argue that the emergence of a rule of the above kind is accompanied by a decrease in attention to the non-diagnostic features. And this is why a purely statistical approach to categorisation falls short: it has no ability to weight various features according to their importance to the categorisation task. The rule-network, on the other hand, constantly monitors the statistical network and provides a means of achieving that weighting.

Our model is designed only to learn positive diagnostic rules, e.g. "if X has a beak, X is a bird". One way to teach the system negative diagnostic rules, such as "if X is under 18 X can't vote", would simply be to define explicitly negative categories (in this case, "can't vote"). One potentially more serious limitation is that the model can verify only the conditional statements (if *p*, then *q*), and not their contrapositive (if ~q, then ~*p*). In other words, the system will learn, "If it has trunk, it is an elephant", but cannot check that "if it is not an elephant, it does not have a trunk". Since, technically speaking, verifying the rule requires checking the validity not only of the conditional, but of its contrapositive, our system is not doing traditional rule-learning. However, in terms of the evolution of human cognition, the type of rule learned by the present system, however incomplete from the standpoint of Aristolean logic, would still have provided animals with a significant adaptive advantage over those lacking this mechanism. We therefore suggest that our mechanism is a plausible account of the way in which humans attain at least a subset of the rules they acquire.

The importance of rules

There is evidence that young infants perform categorisation of cats and dogs in a purely bottom-up manner, basing their category discrimination on the statistical distributions of the perceptual characteristics of the two categories (Mareschal, Quinn, & French, 2002; French, Mareschal, Mermillod, & Quinn, 2004). On the basis of this research, it seems likely that, under the age of 3-4 months, infants do not learn rules underlying category structure. Rather, the data seem to indicate that they perform categorisation using a strategy that does not differentiate between features that are simply correlated with category membership and features whose presence or absence can be used to diagnose category membership.

There are at least two ways in which such a strategy might be disadvantageous. First, attending to all perceptual features of stimuli, when the application of a simple rule would suffice for categorization, squanders cognitive resources. Second, and more importantly, a purely bottom-up strategy can lead to misclassification of certain types of novel stimuli.

Consider a person who wishes to sort shirts according to brand. Many features can be used for this sorting, including the quality of the fabric, the quality of the sewing, the number and type of buttons, the presence/ absence of a collar, etc. But one day he realizes that if there is a little green crocodile anywhere on the shirt, it is a "Lacoste" shirt. Henceforth, he can identify Lacoste shirts without paying any attention whatsoever to the other features. He has extracted a rule: IF *green crocodile*, THEN *Lacoste*. One day he sees a shirt that unlike any he has seen before: it is made of leather, has pearl buttons and leaves the wearer's navel exposed. But it has a little green crocodile over the left breast. His rule allows him to ignore the other features of the shirt and conclude, albeit with some surprise, that it is a Lacoste shirt.

In short, to go from attending to all features to attending to only a small subset of category-specific diagnostic features, one must learn which features to ignore. During the acquisition of the rule, features associated with several categories must "drop out" of the representation in the rule network. This elimination of features as diagnostic for categorisation signals the emergence of a rule.

Operation of the model

The essence of the present model is the tandem operation of a statistical-learning (Kohonen) network and a rule-extracting network (driven by competitive learning) that continuously monitors the state of the statistical-learning network. The overarching principle of the rule-extracting network is as follows. If a particular input (i.e., feature) unit in the Kohonen network has a high-valued weight connecting it to only one category output node, and small weights to all other category output nodes, then that feature is a defining feature for that category, one which we will refer to as a "diagnostic" feature. For example, in Figure 1 the weight between *beak* and *bird* will become large during training, while the weights between *beak* and any other category node will remain small (because only birds have beaks).

The rule-network consists of a copy of the original Kohonen network in which competition between the weights emanating from each feature node determines which feature nodes are important for categorisation. When a particular feature (e.g., *eyes* in Fig. 1) is shared by a number of categories, the competitive-learning process pushes down the values of all of the weights emanating from the *eyes* feature unit in the rule network, so that *eyes* is not a diagnostic feature for any particular category.

The category response of the network to a given novel stimulus is a linear combination of the output of the statistical (Kohonen) network and the rule network.

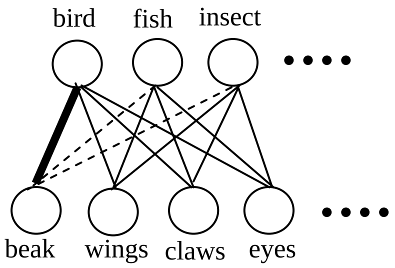

Figure 1. Any animal for which the first feature (*beak*) is active is a *bird*. In the Kohonen weights, the weight between *beak* and *bird* is large, while the *beak-fish*, *beak-insect*, *beak-automobile*, etc. weights are small. This is what the rule-network notices.

Implementation details of the model

The Kohonen network used in our model is a two-layered network with perceptual feature nodes on input and category nodes on output. During learning, neighbouring regions of the output layer are trained to represent stimuli with similar perceptual features, so that representations of similar stimuli cluster together. Thus, if stimuli within a category share many perceptual features, they are "classified" by the Kohonen network as belonging to the same category. The network is implemented using leaky integrators and interneurons to provide neurobiological plausibility, since it has been argued that this type of network exists in visual cortex Kohonen (1993).

Statistical-Learning Network

Kohonen networks are designed to model the type of neural processing that occurs in mammalian cortex. The Kohonen network in the present model comprises a one-dimensional array of processing units that receives stimulus inputs from the input layer and implements lateral excitation and inhibition between neighbouring units (Figure 2). The weights from input units (feature units) to output units (or category units) are trained by the successive presentation of a number of stimulus inputs; units' weights are incrementally adapted on each presentation via a Hebb-type learning rule. This results in an automatic mapping of stimulus inputs onto a set of representations that possess the same topological order as the stimuli, that is, similar stimuli are represented in neighbouring locations on the output layer.

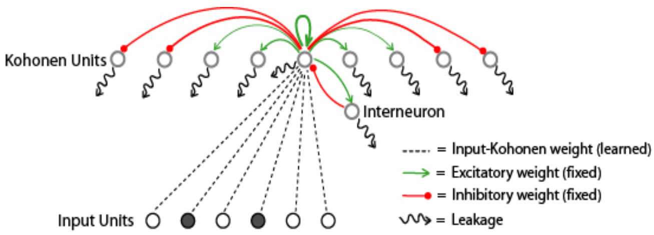

Figure 2: Statistical Learning Component of the Model. For clarity, only 6 units are shown in the input layer; there are 10 input units in the model. All output units are coupled with an interneuron and have lateral connections to all other output nodes. All input units are connected to all output units.

We have attempted to implement the Kohonen network of the present model in a biologically plausible manner; many of the neuron-like properties are illustrated in Figure 2. Lateral interactions are implemented directly between individual output units. Each unit in the output layer receives input from the input layer, from an inhibitory interneuron, and from other collateral units. The activations of units in the output layer are then calculated iteratively and simultaneously, so that each unit's activation evolves according to the input it receives from other units, whose activations are simultaneously being adjusted. Both output units and interneurons are subject to a degree of activation leakage. This was implemented as a set of non-linear differential equations similar to those described in Kohonen (1993).

Rule Network

While other algorithms have been developed (e.g., Thomas, van Hulle, & Vogels, 2000) for determining the relative importance of the weights in a Kohonen network, one of the aims of our model was to implement the rule-network with structures and mechanisms that could conceivably arise in the cortex. Thus, the overarching idea of this network — comparison of (by means of competition between) the synaptic weights in a copy of the statistical-learning network — was implemented by introducing a set of rule units whose activations could be used to implement this competition (see Figure 3). The rule network consists of a copy of the weights of the Kohonen network that lead, not to the category nodes of the Kohonen network, but to a set of rule units. (We acknowledge that there is currently little biological evidence suggesting a precise mechanism by which this copy might be made). The copied weights and rule units are organized so that for each input feature of the Kohonen network there is a "column" of nodes in the rule network, i.e. one node for each Kohonen weight emanating from that feature. The competition between weights emanating from a feature of the Kohonen network is thereafter implemented as competition between the activation levels of units in the corresponding column of nodes in the rule network.

The detailed operation of the rule network is illustrated in Figures 3 and 4. In Figure 3, the manner in which the rule network monitors the Kohonen network is shown. Each weight emanating from an input feature of the Kohonen network to the Category (output) layer corresponds to a node in a column of "rule units" in the rule network. This column can be said to contain the set of rules pertaining to that input feature, e.g. "if feature F, then Category Y". As shown in Figure 3, competition is implemented among units in a given column via lateral weights. The activity of the rule units allows the rule network to determine which weights of the Kohonen network – and therefore which of the input features – are influential in activating the various category units.

Crucially, the mechanism of competition within each column is what causes activation levels of non-diagnostic features to be depressed in the rule network. Assume that a given feature in the Kohonen network sends high-valued weights to numerous category nodes (e.g., the *eyes* feature node in Figure 1). This will result in a high level of activation of the numerous nodes in the column of rule nodes associated with that feature in the rule network. Mutual inhibition within this column will then depress the activations of *all* of the nodes in that feature column. The result will be that this feature will not be perceived by the rule network as diagnostic for any particular category.

The competition between the activated rule units is implemented on every trial, thus the system gradually determines which feature(s) are diagnostic for membership of each category. This diagnostic information must not only be averaged over trials and stored, but must also be available for retrieval by the network. Both of these aims are achieved by developing a set of "rule weights" that link the original feature inputs of the Kohonen network to the rule units. The rule network's stored knowledge can thus be retrieved by passing the input activation through these weights. The input features now feed into two networks: the statistical Kohonen network, as before, and the rule network.

The connectivity shown in Figure 3 (the copy of the Kohonen weights providing input to the rule units) is used to determine the activity of the rule units *during training*. The connectivity shown in Figure 4 – the set of 'rule weights' – is used to determine the activity of the rule units *after training*, and hence to determine the output of the rule extracting component of the model when confronted with novel stimuli. The rule weights are learned by a Hebb-type algorithm that depends on both the input unit activations and the rule unit activations.

The competition for activation within each column of rule units is implemented by each rule unit having a recurrent, excitatory link to itself and inhibitory links to all other units in the column. The activation of rule units is determined first by passing activation from input units to rule units, then by iterating the activations of all rule units in the column through the mutually inhibitory lateral weights for a fixed number of cycles.

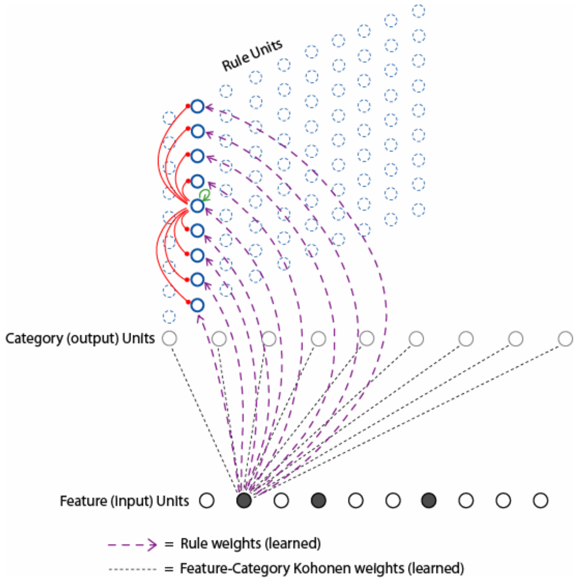

Figure 3: The statistical-learning component of the model is shown at the bottom of the figure and the rule-extracting network is shown above it. The two components share a set of input units. Note the arrangement of the rule units and the connections providing their input (the 'copy' of the Kohonen weights). These connections are employed in determining the activation of the rule units during training, and are instrumental in monitoring the 'knowledge' in the Kohonen network.

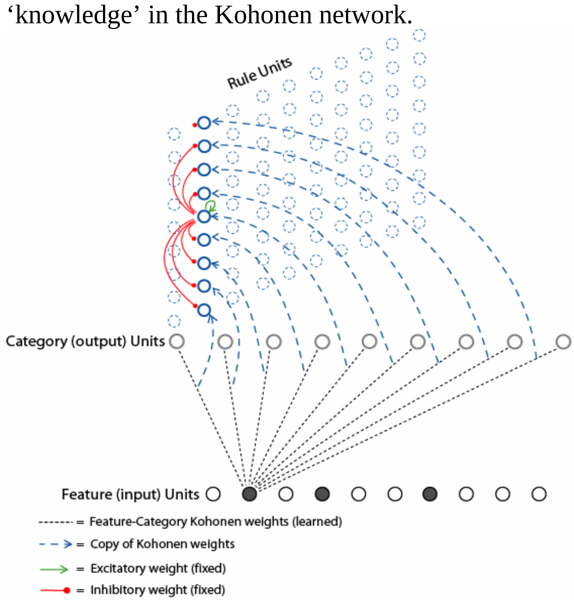

Figure 4: The "rule weights" of the rule network connect the feature units to the rule units. The rule weights are learned via a Hebbian process, which depends on the rule unit activations, which are determined by the connectivity shown in Figure 3.

Input and output

The model is trained with stimuli from three categories of objects. Many exemplars from each category are presented to the Kohonen network. Since, after training, each category becomes associated with a particular region of the Kohonen network output layer, any output unit in this region will be said to "represent" the associated category. (The units in the center of the region are, in general, better representatives of the category than those on the periphery of the region.) The model output can therefore be interpreted as a "choice of category".

During the test phase, the model is presented with a novel stimulus and we consider three different outputs from the system: the response of the statistical learning component alone, the response of the rule network alone, and the linear sum of the responses from both components of the model. For the response of the statistical learning component, we take the most active unit in the output layer of the Kohonen network. To determine the rule-network response, we send the input stimulus activation through the rule weights and sum the activations of the rule units across the columns for each category output node, i.e., there is a *row* of rule units for each category output node. For the "combined" response, the activation of the output units of the Kohonen network is linearly combined with the activation values from the output nodes of the rule network. The greatest combined activation value determines the model's response.

Simulations

Stimuli

Stimuli were represented as an input vector with ten elements (or 'features'). Each feature may be thought of as some real-valued property. All stimuli had two high-valued elements (i.e., features that are present) and eight low-valued elements (i.e., features that are absent). These values differed for each stimuli, but, for example, Category A stimuli always had high values on features 7 and 8 and low values elsewhere. Specifically, each high-valued feature could take a pre-normalisation value in the range 0.6 to 1, while low-valued features varied between 0 and 0.1. All stimuli were normalised. The stimuli were divided into three categories, A, B, and C, as shown in Figure 5. Categories A and B had an overlapping (and thus non diagnostic) feature: 8. Each category was defined by at least one *sufficient* feature.

We trained the model on the three categories of stimuli and then tested it on a novel, but atypical stimulus. This test item was a stimulus that, because of the presence of a diagnostic feature (10), belonged to category C, but also had perceptual overlap with stimuli from categories A and B because of its (non diagnostic) feature (8).

Method

The model was trained by presenting 200 exemplars from each of the three categories. The weights of the Kohonen network and the rule network were updated on every stimulus presentation. After training, the combined network was presented with an exemplar from each of the three categories to ensure that it classified novel elements of each category correctly (it did). Then, to demonstrate that the network's acquisition of rules actually made a difference in its classification behavior, we tested it on an "atypical" test stimulus. This was a stimulus that contained at least one diagnostic feature that meant that it belonged to a certain category, but also included other "distracter" features that were irrelevant for categorisation. The idea was that, once the network had learned the rule associating that diagnostic feature with a particular category, it would ignore the distracter feature(s) and produce a correct classification. On the other hand, the Kohonen network alone would be misled by the distracters and would misclassify the stimulus.

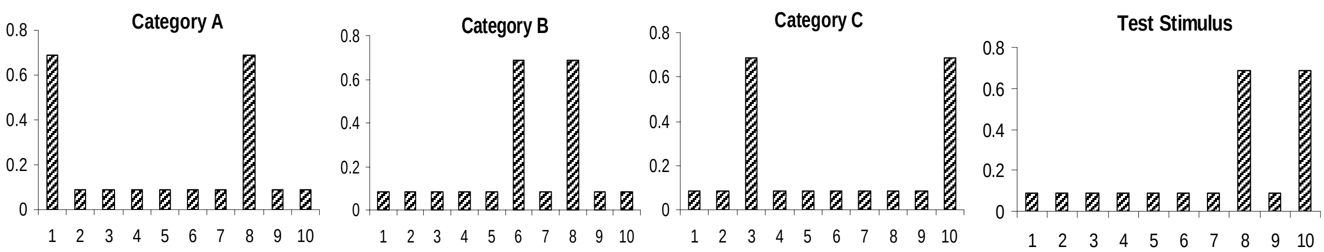

Figure 5. The value of each element in the input vector is shown for a prototypical item from Categories A, B and C, and for the Test Stimulus. Stimulus input vectors are normalised. Note that feature 8 is shared by both Categories A and B and is thus not a diagnostic feature. The test stimulus includes one diagnostic feature (10) for Category C as well as feature (8), which belongs to both Categories A and B.

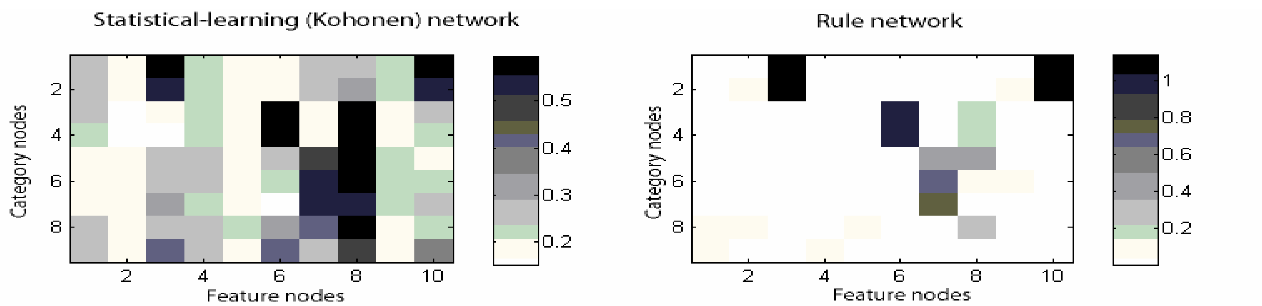

Figure 6. Left: Weight values of the Kohonen network trained on stimuli from categories A, B and C. Right: the rule network has discovered that feature 8 is far less relevant than features 3, 6, 7 and 10 for category determination

Results

As training progressed, exemplars from each of the three categories began to activate consistently the same region of the output layer of the Kohonen network. The diagrams in Figure 6 represent the weight values – from feature nodes to category nodes – of the Kohonen network (on the left) and the rule network (on the right). The pattern of weight values of the Kohonen network shows that Category A items (features 7 and 8 active) are represented by output units 5, 6 and 7), that Category B items (features 6 and 8 active) are represented by output units 3 and 4; and Category C items (features 3 and 10 active) are represented by output units 1 and 2.

Of paramount importance is what happens to feature node 8, a non diagnostic feature shared by items in both Categories A and B. In the Kohonen network, the weights produce — as they should — the activation of category units 3 through 8, the units corresponding to Categories A and B. But when we look at the column of weights for this feature in the rule network, we see that the weights are all low-valued. This has arisen because of the mutual inhibitory competition from the large number of strong feature-to-category weights associated with feature 8 in the Kohonen network. Feature 8 has effectively dropped out of consideration as a diagnostic feature.

The novel test stimulus (Figure 5) has an active feature 10 that makes it a Category C item and also has an active "distracter" feature 8. When this stimulus is presented to the system, the Kohonen network alone classifies it as belonging to Category A, while the rule network alone, as well as the combined rule-and-Kohonen network, classify it correctly as a Category C item.

Conclusion

We have presented a dual-network connectionist model of unsupervised categorisation using two interacting networks: a Kohonen network for extracting statistical information from the input and a competitive-learning network that extracts rule information from the Kohonen network. The addition of the rule network allows the system to correctly categorise novel, but unusual, items that the Kohonen network alone misclassifies.

In addition, preliminary simulations indicate that the model is also able to perform supervised category learning, which leads to an interesting observation. While the model can perform categorisation *with* feedback for stimulus categories with no clustering in stimulus space (e.g. the separable but not clustered categories of Erickson and Kruschke 1998), it can only perform categorisation *without* feedback if the stimuli cluster naturally into categories. This pattern surely echoes human behaviour: in an unsupervised version of Erickson and Kruschke's categorisation task, subjects would not have spontaneously categorised the stimuli according to the experimenter-imposed boundary. Category learning in a natural context generally proceeds with little or no feedback, but, happily, tends to involve categories that are perceptually clustered, at least for living things, making the unsupervised task much easier. This highlights a potentially fundamental difference between artificially constructed, supervised, categorisation tasks, in which categories are not clustered in stimulus space, and the type of category learning behaviour that is exhibited in a natural environment.

Acknowledgments

We acknowledge EC FP6 NEST Grant. 516542. We also thank Denis Mareschal for insightful comments on this work.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning.. *Psych. Rev., 105,* 442-481.

Chalmers, D. (1990). Syntactic transformations on distributed representations. *Connection Sci.,* 2, 53-62.

Erickson, M. A. & Kruschke, J. K. (1998). Rules and Exemplars in Category Learning. *JEP:G***,** 127**,** 107-140.

French, R. M., Mareschal, D., Mermillod, M. & Quinn, P. (2004) The role of bottom-up processing in perceptual categorization by 3- to 4-month old infants: Simulations and data. *JEP:G*, 133, 382-397.

Kohonen (1993). Physiological interpretation of the self-organizing map algorithm. *Neural Networks, 6,* 895-905.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43,* 59-69.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psych. Rev.,* 99**,** 22-44.

Marcus, G. F., Vijayan, S., Bandi Rao, S., and Vishton, P. M. (1999). Rule-learning in seven-month-old infants. *Science*, 283, 77-80.

Mareschal, D., Quinn, P. C., & French, R. M. (2002) Asymmetric interference in 3- to 4-month-olds' sequential category learning. *Cog. Sci.*, 26, 377-389.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition and typicality. *JEP:LMC*, 14(4):700–708

Pinker, S. & Prince, A. (1988) On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*, 73-193.

Posner, M.I. (1986). Empirical studies of prototypes. In Craig, C. (ed.) *Noun Classes and Categorization*. John Benjamins, Philadelphia, 53-61

Reber, A. S., (1967) Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior,* 6, 855-863.

Rosch, E. 1978 'Principles of categorization', in Rosch & Lloyd (eds) *Cognition and Categorization*, NY: Wiley.

Seidenberg, M.S., & McClelland, J.L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96,* 523-568.

Thomas E., Van Hulle M. & Vogels R. (2000). Encoding of categories by non-category specific neurons in the inferior temporal cortex. *J. Cog. Neuro*, 13:190-200.

Wittgenstein, L. (1951). *Philosophical Investigations*. Translation by G.E.M. Anscombe, Oxford: Blackwell, 1963