Int. Conf. on Artificial Neural Networks, ICANN (1992), Brighton, UK

Unsupervised symmetry detection: A network which learns from single examples *

Wolfgang Konen
Christoph von der Malsburg
Institut für Neuroinformatik, Ruhr-Universität Bochum, FRG
Email: wolfgang@neuroinformatik.ruhr-uni-bochum.de

Abstract

Learning problems of higher order, like detecting an unknown symmetry within a pattern, are difficult tasks for most neural networks. Even for small problem sizes a very large number of training examples is needed. We propose a self-organizing network which learns unsupervised to categorize the symmetry of new input patterns according to symmetries already seen. Single examples of a given symmetry are sufficient for learning. The network achieves a classification reliability of 96% percent (3 classes). The underlying mechanism is based on the self-organization of dynamical links which is driven by correlated activity of cell assemblies.

Introduction. Neural networks are widely used for tasks where a high dimensional input pattern has to be classified according to a few possible output classes. They have proven to be very successful in cases where the rules for the input–output mapping are complicated or even not known [1, 2]. Consider for example the famous NetTalk system, a network which learns from examples to translate written English text into the proper phonemes. The identification of the right phoneme depends partly on the actual letter (problem of order one[1]), but partly also on the context, i.e. the letters before and after that letter (problem of order two or higher). Standard neural networks have been especially successful in this hybrid situation.

But what happens, if the problem is of an order strictly higher than one, that is, none of the possible input examples allows classification from looking at only one input cell? As an example one may consider the detection of global symmetries in a 2D-pattern, i. e. mirror symmetry with respect to an arbitrary axis (Fig. 1). Here the interesting information resides only in relations among the local features and not in the features themselves. Many problems in vision or invariant pattern recognition are of this kind.

It has been shown [4] that a standard neural network algorithm (Boltzmann machine) can solve such a problem for small problem sizes (4×4 or 10×10 pixels). This network has 12 hidden unit neurons with initially homogeneous connections to all input cells. Thus it has no a priori knowledge about the learning task. During learning, the hidden units develop weight patterns with the same symmetries as those of the shown examples. However, the number of examples needed to train the network is very large (several 10^4) and grows rapidly with the problem size. It has been suggested in [4] that the order of the problem rather than its size is the appropriate measure for the difficulty of training a network to solve this problem. The

*Supported by grants from the German Ministry for Science and Technology (413-5839-01 IN 101 B/9), and a research grant from the Human Frontier Science Program.

[1] A formal definition of the order of a problem was first given by Minsky and Papert [3].

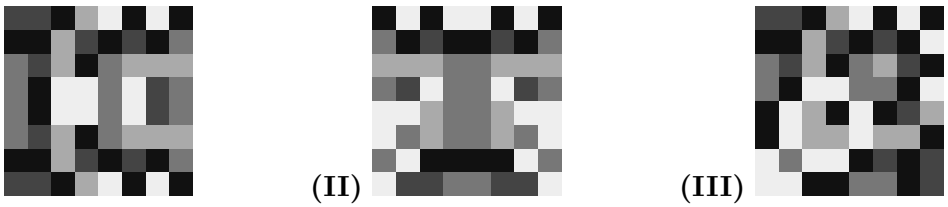

Figure 1: **Symmetrical Pixel Patterns.** Input patterns are arrays of $N \times N$ pixels, here $N = 8$. Each pixel a has one out of 10 possible gray level values F_a. In each input image, pixel values are random, but equal for points symmetrical with respect to one of three axes: **(I)** horizontal, **(II)** vertical, **(III)** diagonal. The system has to solve the task of assigning input patterns to classes according to these symmetries, and to learn this performance from examples.

learning set size problem of artificial neural network algorithms is very striking when compared to our visual system which easily learns new symmetries from only a few examples.[2]

One possible source for the learning set size problem may be the following: Neural network algorithms use the information of the training example to determine at most *one* update of synaptic weights. Such a strategy is well adapted in cases where no a priori information about the internal structure of the training examples is available. It relies only on quantities deducible from the statistics over many examples. However, if the number N of input cells becomes large, the number of statistical moments of order k grows like N^k and it becomes too costly (in terms of training examples) to keep track of all of them. New approaches try to overcome the slow convergence rates of supervised learning by maximizing the mutual information [6].

In the following we will propose another strategy based on a rapid self-organization process which makes use of the topographic constraints of the problem: From a given input example a whole sequence of network activations is generated. Each activation probes a different region of the input and leads to a change of synaptic weights, which in turn influences further activations. In such a way the weights acquire a dynamic behavior with respect to the (static) input pattern. It has been proposed, that such a dynamic behavior plays an important functional role in the brain where short term connectivity changes are controlled by the temporal correlation of nervous signals [7].

Self-organization of dynamic links. Our system consists of two essentially identical neural planes X and Y which express the input pattern in terms of one feature-specific cell per position. Cells in position $a \in X$ and $b \in Y$ have activities x_a and y_b and are connected by a dynamic link $J_{ba} \in [0, 1]$. Initially, all links have the same value $1/N^2$.

The specific self-organizing process which was used in this work to structure the dynamic links (weights) may be described best with the help of Fig. 3 showing some iteration steps of the activation sequence. An iteration step starts with the activation of a large, connected region ("blob" or cell assembly) in layer X (dark circles in Fig. 3).[3] This is achieved with a set of differential equations coupling the cells of X through an interaction kernel $K_d = \exp(-\|d\|^2/2s^2) - \beta$:

$$\dot{x}_a = -\alpha x_a + (K * X)_a + \rho_a, \qquad X_a = \sigma(x_a).$$

where $\sigma(\cdot)$ denotes a sigmoidal output function. The center of the blob region may appear anywhere in layer X, its position being imposed by internal noise ρ which changes randomly from step to step. The functional role of the blob is similar to that of the "searchlight" suggested by Crick [8].

[2]From a computational point of view it should be noted that a large training set size is a non-parallelizable cost, thus enlarging the training time even on an ideal parallel machine [5].

[3]In order to avoid boundary effects, we are assuming wrap-around conditions for the X and Y layers.

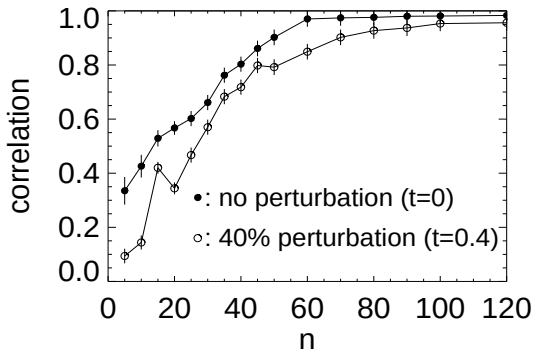

Figure 2: **Mean correlation between pairs of corresponding cells** in layer X and layer Y for a given state of the dynamic links J after n steps in the activation sequence. **(a)** Filled circles: Perfect feature similarity function $T(b,a) \in \{0,1\}$. **(b)** Open circles: All matches $T(b,a) = 1$ are replaced by random values $T(b,a) \in [1-t,1]$, all non-matches $T(b,a) = 0$ by a random $T(b,a) \in [0,t]$, to mimic the effects of noisy feature information. The correlations are robust against these perturbation.

Simultaneously with the blob formation in X, activity flows from X to Y through the dynamic link network, inducing in Y another blob-formation process:

$$\dot{y}_b = -\alpha y_b + (K * Y)_b + \sum_a J_{ba} T(b,a) X_a, \qquad Y_b = \sigma(y_b).$$

The blob will appear in a region of layer Y which receives the strongest activity flow. The activity flow between two cells $a \in X$ and $b \in Y$ is proportional to the link J_{ba} connecting them and proportional to the similarity of the local features assigned to a and b. For the features F_a of Fig. 1 we used a rather simple feature similarity function $T(b,a) = \delta_{F_b,F_a}$. The dynamic link algorithm is by no means specialised to this function. In order to avoid the trivial identity mapping between X and Y, the flow of activity between cells in the same position is suppressed ($T(a,a) = 0$).

When the activated regions in both X and Y have been established, the dynamic links between active cells with similar features are strengthened according to

$$\Delta J_{ba} = \epsilon (J_{ba} + J_0)\, T(b,a)\, Y_b X_a,$$

and all links are normalized by putting first $\sum_a J_{ba} = 1$, then $\sum_b J_{ba} = 1$. Here $J_0 > 0$ is a constant growth parameter.

The new state $\{J_{ba}\}$ of the dynamic link network influences the next iteration step of the activation sequence (starting with another "blob" in X), since it helps to guide the activity flow from X to the corresponding region in Y. As can be seen from the sequence in Fig. 3, the active regions in X and Y do not lie mirror-symmetrically to each other in the initial steps (A), while they are exact mirror-copies of each other in the final step (C).

It is crucial for the self-organizing process that two blobs from different activation steps can have a large overlap: Links of active cells belonging to both blobs can cooperate to form a piecewise topographic mapping from X to Y. It is in this way that the a priori *principle of neighborhood conservation* comes into play: The growth of two links starting from neighboring cells is favored if they terminate also in neighboring cells. Solely with this *a priori* information the network is able to structure its dynamic links in an unsupervised process according to any symmetry presented. The ambiguity of the local feature information is resolved by the global context. Other self-organizing feature maps [9] start with initial weight vectors corresponding to a randomly disordered mapping, whereas here the link network starts in a low-activity state with all-to-all connections, and significant links grow directly in the correct ordering.

In simulations of the process on 8×8 layers it took about 40-50 iteration steps to organize the dynamic links. The degree of organization can be read off by measuring the correlation in

Table 1: **Classification performance** in percent of a network trained from only one example per symmetry class (I)–(III) (cf. Fig. 1) for the cases **(a)** and **(b)** as described in Fig. 2. The overall success rate **(a)** 96%, **(b)** 93% is only weakly affected by local perturbations.

(a)		detected			**(b)**		detected		
		I	II	III			I	II	III
shown	I	95	2	3		I	94	3	3
	II	0	95	5		II	2	95	3
	III	0	3	97		III	5	5	90

activity of cells in X and Y which lie mirror-symetrically to each other. Fig. 2 shows that the self-organization process is robust against local perturbations.

Unsupervised Classification. The correlated activity patterns (CAPs, dark circles in Fig. 3) generated in each activation step may be the input for further processing modules in a more complex system. Here they are used directly for the classification task. The dynamic links themselves are re-structured anew for each new input example from the initial state $J_{ba} = const.$ Permanent changes occur only in a classification network which receives the CAPs as its input.

Here follows a brief description of the specific classification mechanism used in this work: The unsupervised classification into M possible output classes is based on M correlation detectors C_k, $k = 1 \ldots M$, which accumulate for a given input pattern the activity correlations between X and Y. They are enabled to do so by a set of hidden units which have pairs of receptive fields (RFs) in X and Y. Initially, those RFs are random and unspecific. The output of a correlation detector C_k becomes active as soon as its accumulated input exceeds a certain threshold. An inhibitory coupling of the output to all other detectors ensures that at most one C_k is active at a given time (winner-takes-all scheme). Thus, on the first example of an unseen symmetry, one of the detectors, say C_k, will win and suppress the others. This detector then learns to modify its RFs to respond better to the CAPs encountered. If, later on, an input example of the same symmetry has to be classified, it will generate the same sort of CAPs, and the detector C_k will become active again and signal the symmetry already learnt. If, on the other hand, a new symmetry is shown to the system, this will generate a different sort of CAPs and one of the so far untrained $C_l, l \neq k$, will become active and learn the new symmetry. In this way, only the symmetries actually seen are learned and they can be retrieved in later examples (Tab. 1).

Conclusion. The scaling behavior of neural networks learning without any *a priori* structure is known to be rather poor, especially for problems of higher order. We proposed here an unsupervised algorithm based on the self-organization of dynamic links [7] which may overcome this difficulty. As a very general *a priori* structure the algorithm follows the principle of neighborhood conservation. The adaptive behavior of the dynamic links has a twofold purpose: *(i)* through the self-organizing iterative process a large amount of information can be extracted from a given example, thus making it possible to learn from single examples; *(ii)* the dynamic links can represent higher order relations among the input data and can – together with the activation dynamics – generate very specific sets of correlated activity patterns (CAPs) which are highly robust against local pertubations of the input pattern.

Our system is meant to be a simple conceptual sketch and is not intended to serve directly as a model of biological systems. Of central importance to our system, however, is the encoding of significant relations with the help of temporal[4] signal correlations. Candidate signal correlations of an appropriate nature have already been observed in visual cortex [10, 11]. The model may thus shed some light onto the question how correlated activity in neural signals can play a useful and efficient functional role in information processing.

[4]In our model the unit of "time" is an iteration step in the activation sequence.

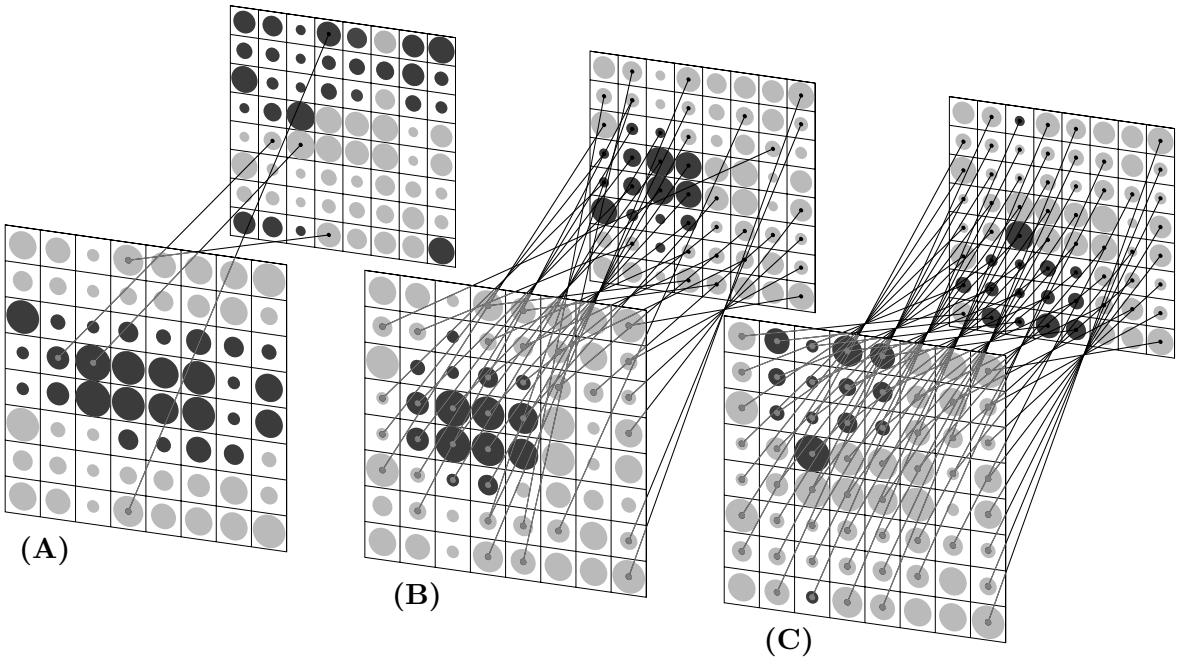

Figure 3: **Self-organized formation of dynamical links.** The figures **(A)**–**(C)** show the layers X (in front) and Y (in the rear) in different activation states generated from a single input pattern. The input pattern imposed to both layers is of symmetry class **I** (cf. Fig. 1) and features are represented here by the diameter of the different circles. The self-organization process consists of a sequence of activations, each of them activating large, overlapping regions in the layers (dark circles). The figure shows the network state after **(A)** 15, **(B)** 50, **(C)** 80 activations. Links $J_{ba} \in [0, 1]$ grow between cells which are active simultaneously and have similar features. Only links with $J_{ba} \geq 0.4$ are shown in the figure.

Acknowledgements. We would like to thank L. Wiskott and S. Tölg for helpful discussion and critical reading of the manuscript.

References

[1] T.J. Sejnowski and C.R. Rosenberg. *Complex Systems*, 1:145, 1987.

[2] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning representations by backpropagating errors. *Nature*, 323:533–536, 1986.

[3] M. Minsky and S. Papert. *Perceptrons*. MIT Press, Cambridge, 1969.

[4] T.J. Sejnowski, P.K. Kienker, and G.E. Hinton. Learning symmetry groups with hidden units: Beyond the perceptron. *Physica*, 22D:260–275, 1986.

[5] S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4:1–58, 1992.

[6] S. Becker and G. Hinton. Self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, 355:161–163, 1992.

[7] C. v. d. Malsburg. The correlation theory of brain function. Internal Report 81-2, Dept. of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, 1981.

[8] F. Crick. Function of the thalamic reticular complex: the searchlight hypothesis. *Proc. of the Nat. Acad. of Sciences*, 81:4586 – 4590, 1984.

[9] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.

[10] R. Eckhorn, R. Bauer, W. Jordan, M. Brosch, W. Kruse, M. Munk, and H. Reitboeck. Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics*, 60:121, 1988.

[11] C. M. Gray, P. König, A. K. Engel, and W. Singer. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338:334–337, 1989.