

HIGHER ORDER ADAPTIVE NETWORKS -
SOME ASPECTS OF MULTI-CLASS AND FEED-BACK SYSTEMS

T.J. Stonham, B.A. Wilkie and L. Masih
Department of Electrical and Electronic Engineering
Brunel University,
Uxbridge, Middx. UB8 3PH

ABSTRACT

Single layer networks of adaptive logic functions have been used extensively for pattern recognition. They do, however, have limitations in as far as the output at a given time 't' is a function of the input at that time only. The system is equivalent to a conventional combinational logic circuit. If some form of feed-back is introduced, the output becomes dependent not only on the current input pattern, but also on all previous patterns input to the system. The network then has sequential properties. Two forms of higher order system are presented in this paper. The histogram of the output responses is recognised by a second layer network. This achieves an effective increase in discrimination and can correct substitution errors on the first layer. The second system involves feed-back of a pattern generated at the output of the net. These output patterns are prototypes which the system has been trained to recognise. The feed-back leads to convergence in pattern space and a consequent reduction in the variability within a given pattern class. This behaviour has implications in the area of pattern recognition, associative memory and image recall systems.

I INTRODUCTION

A. Single-layered Systems

The single-layered adaptive logic network has found extensive application in the areas of signal and image processing and pattern recognition (Fig.1). The methodology has evolved from two principal sources, neural modelling and statistical methods, and focusses on the use of a combinational logic function as a simple neuron-like cell which fires, i.e. outputs logical one, when a combination of n binary inputs is applied to the function. Practical networks can contain in excess of a hundred thousand functions, each having typically eight inputs, although this number can vary between two and sixteen. The inputs to each function are obtained by sampling an input pattern which can be a binarised television resolution picture held in a framestore. The sampling points for each function may be assigned by an arbitrary but reproducible mapping. However, optimisation strategies are available to tailor the sampling to give improved performances for specific applications (Ref.1).

Logic functions are set up during a training phase where patterns which are typical examples of their class are input to the system. One net is required for each class in the single layer or first order system. Each function samples n points from the binarised input pattern and the n-tuple sub-pattern becomes a term of that logic function. During initial training the functions are frequently up-dated but stabilise after a representative set of patterns has been seen. The precise training requirements vary according to the particular application; however, in a real time system a few tens of seconds, or video data amounting to between two and four hundred images, are usually adequate. On stimulating the network with an unknown pattern, the logic functions are driven by the n-tuple samples and

respond with 0 or 1 according to whether the current sample has been seen during the training. The overall network response is the arithmetical summation of all the function outputs. In a practical system each class has its own network. The selectivity of the network is controlled by the sample size n. The response can be extrapolated from the simple case using one training and one test pattern, where the most likely response of a discriminator r_j is given in equation 1

$$r_j = \left(\frac{k}{n}\right) \times \left(\frac{k-H}{k}\right)^n \quad (1)$$

where k is the dimensionality of the input, n is the n-tuple sample size and H the hamming distance between training and test pattern.

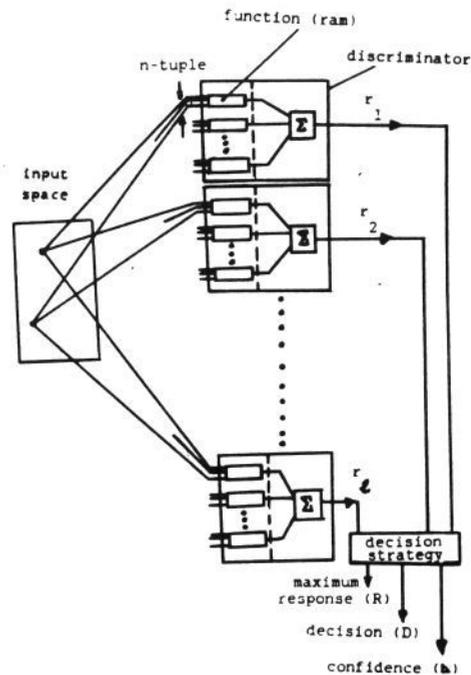


Fig.1 Single layer adaptive logic network.

B. Higher Order Systems

The single-layer (SLN) system is a first order adaptive logic structure. It is combinational and the output $r_j(t)$ is determined solely by the current input pattern I_t

$$r_j(t) = F [I_t]$$

As such, it is unable to recognise sequences of patterns. Other limitations arise from the use of the maximum response alone to define the recognition category. In a system of 'L' single-layer networks (Ref.2) where each

individual net is being trained on a separate class, the decision D is given by

$$D \rightarrow m_l \quad (2)$$

where $r_m = \max_{i=1}^l r_i$

and $\underline{R} = (r_1 r_2 \dots r_l)$

The system response vector \underline{R} is a transformation of the input image

$$\underline{R} = T\underline{I}$$

Basing the classification on a single element of \underline{R} results in considerable information being discarded. A more serious situation, however, arises with a j th class pattern failing to stimulate a maximum response from the j th net, such that

$$I_j \rightarrow D = j \quad 3$$

but $\max_{i=1}^l r_i = r_j$

A misclassification of the input pattern then arises.

A display of the response vector \underline{R} as a bar graph is termed the response profile of the input with respect of an l -class classifier. Experience with applications ranging from industrial part sorting to face recognition has shown the overall response profiles remain consistently similar to the human observer for members of a given input pattern set. Small local variations can, however, lead to an incorrect classification as defined in equation 2, without significantly altering the overall profile.

A two layer recogniser has been evaluated with the second layer operating on the complete profile of the first layer system and recognises it as a transformed image of the original input. The salient features are:

- a) local errors in the first layer system can be corrected,
- b) the first or primary net can act as a pattern transformation for patterns outside the l -classes. Secondary recognition can then enable discrimination on a much reduced image with considerable savings in hardware,
- c) cascaded networks effectively magnify the n -tuple size without incurring the hardware cost C of increasing sample size n in the primary net. The cost is an exponential function of the n -tuple size in a single layer system as given by equation 4.

$$C = \frac{k}{n} \cdot 2^n \text{ bits per discriminator} \quad (4)$$

The other higher-order system presented in this paper involves the generation of a binary image at the output of the network, which can then be fed back to the input space and the system cycled. The summation process in the single-layer net is replaced with an image store and individual logic functions can be trained to either logical 0 or 1 in order to output the desired image. The pattern generated at the output and the composite image reapplied to the net. The aim of this feedback investigation is to compare open-loop and closed-loop systems for clustering and image reconstruction properties. Positive clustering occurs if the diversity of the output states in a feedback system is significantly less than in a test set of input patterns. Image reconstruction involves the regeneration of a complete image from partial or distorted data. Both objectives are of fundamental importance to pattern recognition.

II SOME PERFORMANCE LIMITATION OF SLNs

SLNs provide excellent recognition performance for non-random type patterns which contain high detail. However, experience has shown that images with either a high 'black' or a high 'white' content can provide inadequate and confused decisions. Consequently, it is recommended that images presented to an n -tuple system should be pre-processed to provide an equal black/white content. If, however, a binary image of an object has a bit density very much less than 50%, substantial amounts of noise will be introduced in the processed image if thresholding to produce equal black/white content is applied. Many industrial pattern recognition problems fall into this category. However, in this paper familiar patterns will be used. Playing cards are an example of patterns where the detail can be as low as 2% of the image, and banknotes, where a high degree of detail is present.

A. Data Base

The data used in this evaluation were images of numbered playing cards and bank notes. Numbered (non-face) playing cards were chosen because the resultant images are of low complexity with very large areas of white and, between cards of adjacent orders of magnitude, near match patterns can occur. Bank notes were chosen because their images contain relatively high detail where, generally, a single WISARD* system would still provide a good performance.

In practice, for the playing card images, the relevant second choice discriminator responses are not necessarily reciprocated, the reasons for which are outlined in the following simple example.

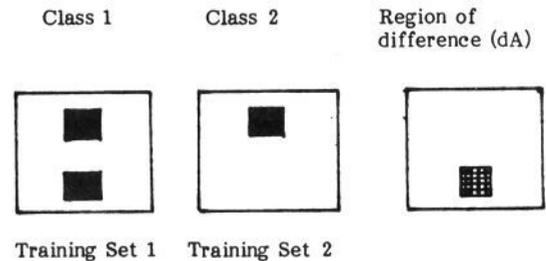


Fig.2 Training set patterns which can produce asymmetrical response profiles.

With reference to Figure 2, if classes 1 and 2 are trained without positional variance, the response differences (R) are given by the expression:

$$R \approx \left(\frac{dA}{A}\right)^n$$

where dA = region of area difference
 A = area of commonality between the two images
 and n = size of n -tuple.

However, if positional variations occur throughout the data base, then the above expression no longer holds as dA must be specified for every combination of pairs of patterns.

For patterns with positional variance the following analysis can be applied to the dichotomy of Classes 1 and 2 in Figure 2.

* WISARD - a high resolution pattern recognition system based on the n -tuple methodology and single layer net architecture.

Case (1) Training on Class 1 with no positional variance.

$$P(k) = (A-dA)^{n-1} dA$$

where $P(k)$ is the probability of a useful function with the single layer net leading to a confidence

$$\Delta = \frac{n}{k} P(k)$$

Case (2) Training on Class 1 with positional variance. The overall region of difference dA is

$$dA' = T_1 dA_1 \cap T_2 dA_2 \cap \dots \cap T_t dA_t$$

where $T_i dA_i$ is the difference region of the i^{th} training patterns with respect to Class 2 patterns.

Hence with positional variance, $dA' \rightarrow 0$

$$P(k) \rightarrow 0$$

Consequently 'Class 2' \in 'Class 1'

i.e. if Class 2 is presented to the system, a maximum response will also be obtained from the Class 1 detector.

Case (3) Training on Class 2 with positional variance.

$$P(k) = (A-dA)^{n-1} dA$$

Positional variance does not introduce further change in the dA region as no information is introduced into this region of the image space from Class 2 patterns.

$$\begin{aligned} \text{Hence } dA' &\rightarrow 0 \\ \text{and } P(k) &> 0 \\ \text{and hence } \Delta &> 0 \end{aligned}$$

Therefore on classification 'Class 1' \notin 'Class 2'

i.e. if Class 1 is presented to the system, a maximum response will only be obtained from discriminator 1. Therefore, the relative responses of the two discriminators to the two patterns are asymmetrical.

A Class 1 pattern detector will accept both Class 1 and Class 2 type patterns but a Class 2 detector will only accept its own patterns.

In practice, errors can arise because although Class 2 patterns would, in theory, evoke maximum responses from both pattern detectors, the presence of random noise can reduce either response with equal probability and if

$$r_1 > r_2$$

then 'Class 2' \rightarrow 'Class 1'.

The subsequent recognition of the overall response profile with a second layer network offers the potential to correct these errors.

III EXPERIMENTAL METHODOLOGY

For all experimental sets, two cascaded WISARD systems were used and these were the original Brunel WISARD and the later, commercial, CRS WISARD. In this text, for convenience, the Brunel system is defined as W1 and the CRS system as W2. Also, the histograms generated by W1 are termed H1 and those generated by W2 are designated H2.

In essence, for all experiments, 2 individual training sets were implemented using the following procedures:

1. Training phase 1. W1 trained on image.
2. Training phase 2. With W1 classifying image and generating H1, W2 trained on H1.
3. Classification. W1 classified the images and W2 classified H1 and generated H2.

For the playing cards, in each WISARD system, the first discriminator (Discriminator 0) was trained on 'Aces', the second discriminator (Discriminator 1) on 'twos' and so on until the last, the eighth, discriminator (Discriminator 7) was trained on 'eights'. The cards were assumed to be vertical but were otherwise unconstrained within the input window.

A similar procedure was adopted for the bank note evaluation in that, in each WISARD system, a single discriminator was allocated to each bank note. Also, the notes were assumed to be horizontal but otherwise unconstrained within the input window.

IV EXPERIMENTAL RESULTS

Some of the experimental results are shown in Figures 3a, 3b, 4a and 4b. Figure 3a shows the results obtained for test images where confusion can occur between near match patterns. Interpretation of Figure 3a is as follows.

1. The top histogram (H1) is that generated by W1 from the image shown in the lower windowed bright up area.
2. Below histogram H1 is histogram H2 which is generated by W2 from the windowed section of histogram H1.
3. The order of the bars in both histograms is from top to bottom. That is, the top bar of H1 represents the response amplitude of the 1st discriminator of W1 and the 7th bar down represents the 7th discriminator response. Similarly, the 3rd bar down on H1 represents the amplitude response of the 3rd discriminator generated by W2.
4. The maximum amplitudes of histograms H1 and H2 represent maximum responses of 100% and on H1 each vertical division represents an incremental change of 10%.

Figure 3b shows the results obtained when the primary system was trained with a low 2-tuple size and the secondary system was trained with an 8-tuple size. In effect, the secondary system 'magnifies' the confidence levels of the primary system. Consequently, with careful choice of system parameters, such a cascaded WISARD configuration provides a pseudo high n-tuple size system.

Interpretation of Figure 3b is the same as that of Figure 3a.

Figures 4a and 4b show the results obtained for a cascaded WISARD system classifying with and without added noise.

Interpretation of Figures 4a and 4b is as for Figure 3.

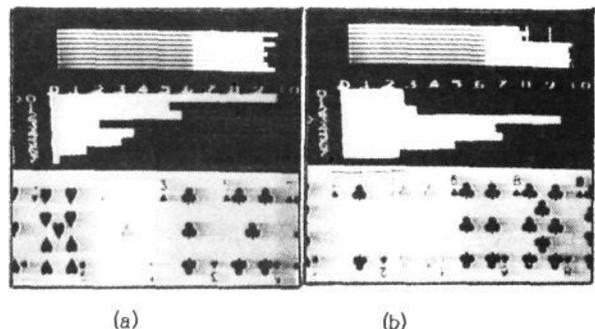
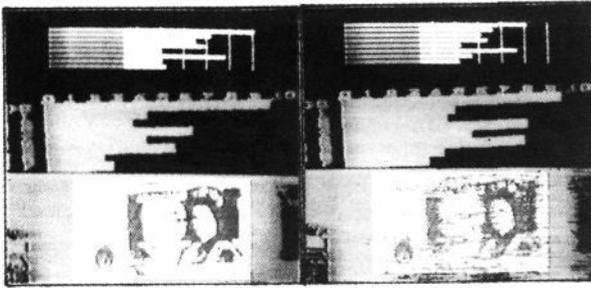


Fig.3 Cascaded system typical responses.

Array size W1 128 x 180 and n = 2
W2 192 x 80 and n = 8



(a) No Added Noise (b) With Added Noise

Fig.4 Effect of Noise on Cascaded System.

Array size W1 384 x 182 and n = 14
W2 272 x 96 and n = 8

V INVESTIGATION INTO PATTERN FEEDBACK

In previous work in pattern feedback (Ref.3) training is achieved by driving the teach terminals with the input image (or fixed model), Figure 5. The changeability of the teach patterns during training (in multi-class discrimination) would imply loss of information in the net as each addressed location may be overwritten by the current input pattern. This results in strong dominance of the last pattern seen during training.

'Ageing' processes have been proposed with progressive reduction in the connectivity of the net.

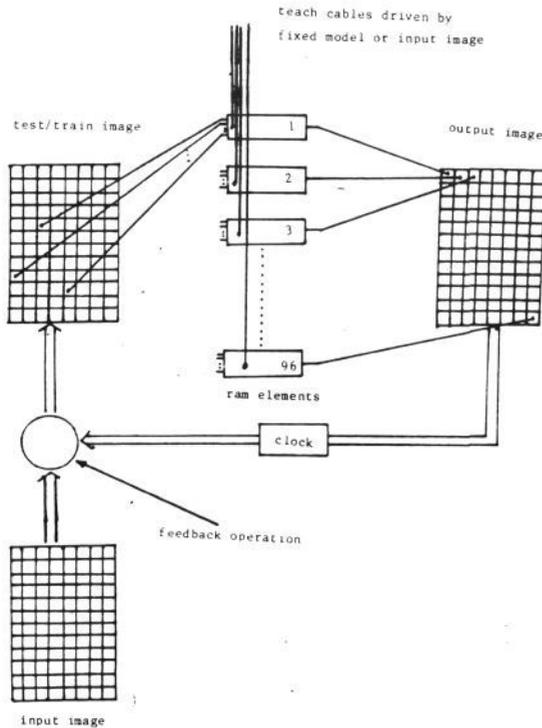


Fig.5 Pattern Feedback Architecture

This 'ageing' process minimises the effects of spurious images and reduces dominance of the last seen pattern. To achieve 'ageing', the training set is applied to the net several times whilst the number of active functions is reduced exponentially. The training process is thus terminated when there are zero functions active, i.e. no

more functions connected to the input. The overall behaviour of the net is still dependent upon the order in which patterns are presented during teaching.

A technique based on frequency of occurrence of n-tuples is proposed as an alternative to the ageing strategy. In this case a count is kept indicating the number of times the function terms are assigned to zeros (-ve count) and ones (+ve count). Once all the training patterns have been taught to the net, then the frequencies are used to set the locations within the functions. Using this technique we can overcome undesirable effects of spurious images and the dominance phenomenon during training.

The other important investigation carried out was to study the effect of varying the type of feedback transformation on the overall performance of the system. Functions including OR, AND and other stochastic operators have been used to control the feedback loop. Results showed strong dependency between the type of transformation used and first order statistical properties of the images in the training set.

A. One Class Performance

The performance of the net can be assessed by the degree of clustering achieved on the incoming test patterns. Clustering is the reduction in hamming distance between the test patterns and the final stable states reached by the net. The extent of clustering was found to depend on system parameters, such as n-tuple size, and the size of the training set.

Figure 6 shows the rate of convergence (final/initial state hamming distance from model) of the net when tested with 100 examples of images belonging to the taught class and those belonging to different classes. The teach cables are driven by a fixed model during the teaching stage.

The results of this experiment show that the net will converge to a far greater extent when tested with examples (not in the training set) of taught numerals than when tested with an unseen class of numerals. All tests were carried out on a data base of machine printed alphanumerics.

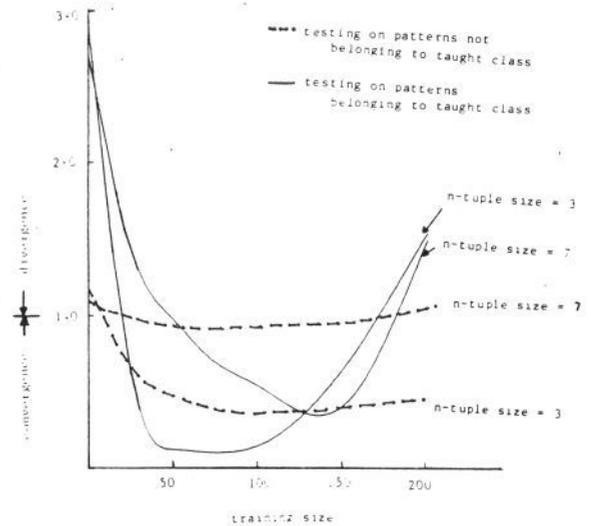


Fig.6 Rate of convergence with varying training and n-tuple size.

B. Multiple Class Performance

The nets behaviour in a multiclass environment was also investigated. The net was trained on three classes of data, namely examples of the numerals 0', 2', and 3'. The fixed models selected were as shown in Figure 7.

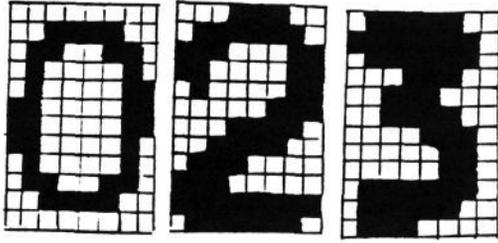


Fig.7 Models selected for driving teach cables.

The net was trained on 400 examples of each class of data and then tested on 100 (not belonging to the training set) of each class, with an n-tuple size of 8. A measure of the intra-class mean and standard deviation in the test and final states was calculated. The diversity of the input patterns was seen to be much greater than that of the stable pattern (as shown in Fig.8) in state space. The net shows capability of switching between these final state space regions under the influence of various stimuli.

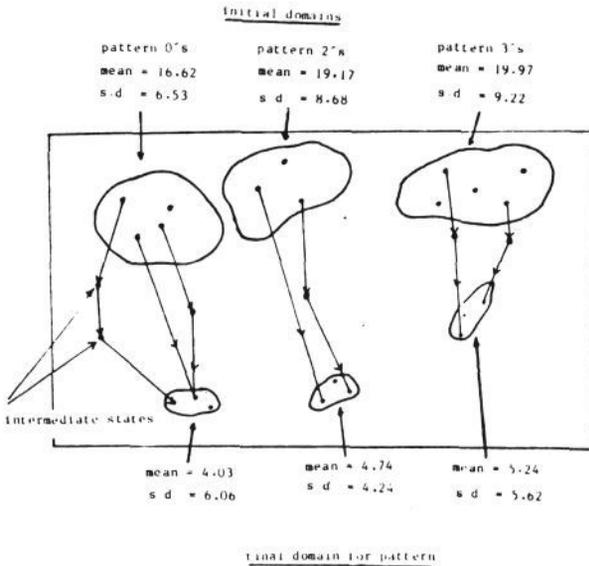


Fig.8 Diversity of patterns in state space regions

VI CONCLUSIONS

Of fundamental importance is that, with low definition images, secondary recognition of the primary response profile corrects initial errors. Also, this correcting effect can be achieved with the use of low n-tuple sizes and secondary image resolutions.

For high definition images, the secondary responses' confidence levels are little improved; however, with suitable training, these responses, compared with the primary responses are unaffected by large positional movement and added noise.

Therefore, sufficient information is contained with the primary response profile for use as a parameter for secondary recognition. This offers a potential for error correction and enhanced performance using profile feedback.

In feedback networks generating output images, a frequency of occurrence measure at each n-tuple site has advantages over a binary system. The system behaviour is no longer dependent on the order of the training patterns.

The investigation has revealed the dependency of the feedback operation on the statistical properties of the training data, in particular the pattern density.

The feedback net is able to discriminate between classes in both a single and multiple class processor. In the single class case the convergence rate is seen to be far less if the stimuli does not belong to the taught class. However, with multiple classes, the net is able to switch between regions in state space, which define the individual classes of data.

A feedback n-tuple system which generates images on the output can reduce the diversity within given pattern classes. Improved pattern labelling can be achieved by operating on the output image of a feedback system.

REFERENCES

1. Patel, M., 'Self Optimisation in Adaptive Systems'. Ph.D. Thesis, Brunel University, 1986.
2. Aleksander, I., 'Emergent Intelligent Properties of Progressively Structured Pattern Recognition Nets'. *Pattern Recognition Letters*, 1983, Vol.1, p.724.
3. Fairhurst, M.C., 'Natural Pattern Clustering in Digital Learning Nets'. *Electronic Letters*, 1971, Vol.7, p.724.

