approximate surface impedance Z_{Sn} of the nth patch is obtained by applying eqn. 4 and normalising to the width of the original region $t/2N$:

$$Z_S^n = j\sqrt{\frac{j\omega\mu\sigma}{w_n\sigma}}\frac{\mathbf{J}_0(j\sqrt{j\omega\mu\sigma}h_n)}{\mathbf{J}_1(j\sqrt{j\omega\mu\sigma}h_n)}\frac{t}{2N} \qquad (5)$$

This expression is the desired approximation for the surface impedance as a function of position on the corners of the conductor.

Results and conclusions: In actual experiment, it is the total internal impedance per unit length of the conductor Z_{tot} rather than the surface impedance, that can be measured. The total internal impedance is the parallel sum of the surface impedances of the various regions, given by

$$Z_{tot} = \left\{2\left[\frac{Z_S^{plate}}{(W-t)}\right]^{-1} + 8\sum_{n=0}^{N-1}\left[\frac{Z_S^n}{(t/2N)}\right]^{-1}\right\}^{-1} \qquad (6)$$

To check the validity of our approximation, we consider the total internal impedance of a square conductor, which should be the most severe test of our technique. Fig. 3 shows a comparison between eqn. 6 and the experimental results of Haefner [6], for square conductors. These data have also been used as a test case in [7]. As shown, even for a very small number of segments, the agreement is excellent. Using only one segment, the error is not worse than 5% at any frequency, and for four segments the error is never more than 0.2%.

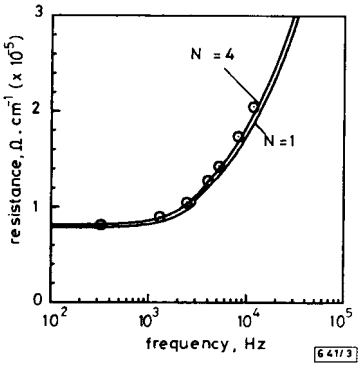

Fig. 3 *Comparison of calculated and measured total internal resistance of square metal bar*

⊙ measured data from [6]
——— eqn. 6

In summary, we have developed a new approximation for the frequency-dependent surface impedance of thick rectangular bars. The expressions are easily used, and are numerically efficient. These results should be useful in the calculation of conductor loss in planar transmission lines and interconnects, providing an accurate transition from DC-like behaviour to skin-depth limited behaviour.

Acknowledgments: This work was sponsored by the Advanced Research Projects Agency Application Specific Electronic Module program.

30 September 1993
Electronics Letters Online No:- 19931375

E. Tuncer and D.P. Neikirk (*Electrical and Computer Engineering Department, University of Texas at Austin, Austin, Texas 78712, USA*)

References

1 ISLAM, M.S, TUNCER, E., and NEIKIRK, D.P.: 'Calculation of conductor loss in coplanar waveguide using conformal mapping', *Electron. Lett.*, 1993, **29**, pp. 1189–1191
2 ISLAM, M.S., TUNCER, E., and NEIKIRK, D.P.: 'Accurate quasi-static model for conductor loss in coplanar waveguide'. IEEE MTT-S Int. Microwave Symp. Dig., 1993, pp. 959–962
3 KENNELLY, A.E., LAWS, F.A., and PIERCE, P.H.: 'Experimental researches on skin effect in conductors', *Trans. AIEE*, 1915, **34**, pp. 1953–2018
4 RAMO, S., WHINNERY, J.R., and DUZER, T.V.: 'Fields and waves in communication electronics' (Wiley, New York, 1984), pp.261–263
5 TUNCER, E., and NEIKIRK, D.P.: 'Highly accurate quasi-static modelling of microstrip lines over lossy substrates', *IEEE Microw. & Guided Wave Lett.*, 1992, **2**, pp. 409–411
6 HAEFNER, S.J.: 'Alternating-current resistance of rectangular conductors', *Proc. IRE*, 1937, **25**, pp. 434–447
7 TSUK, M.J., and KONG, J.A.: 'A hybrid method for the calculation of the resistance and inductance of transmission lines with arbitary cross sections', *IEEE Trans.*, 1991, **MTT-39**, pp. 1338–1347

Design of Hopfield-type associative memory with maximal basin of attraction

J.-Y. Chang and C.-C. Wu

Indexing terms: Associative memories, Neural networks

A method is described for enlarging as much as possible the basin of attraction of a Hopfield-type associative memory. The proposed learning rule, a minimum-overlap learning algorithm that includes a threshold parameter, enables a Hopfield-type associative memory to be designed so that the memory will have a maximal basin of attraction. A technique that diminishes the effect of the threshold on the minimum-overlap learning algorithm is devised. Simulation results show that the basin of attraction constructed by the proposed method is indeed larger than that constructed by several well known methods.

Introduction: The Hopfield associative memory (AM) [1] provides an associative search for storing stimulus-response pair (X^m, Y^m). The AM is now a favoured research subject, for it can play a key role in intelligent systems and pattern recognition applications [2, 3]. Assume that we have a set of M associated pairs of patterns, $\{(X^m, Y^m)\}_{m=1}^M$, where $X^m \in \{-1,1\}$ is the key vector and $Y^m \in \{-1,1\}$ is the recollection vector. We need to form a memory weight matrix W that produces Y^m as its output when the distorted version of X^m is its input; i.e.

$$Y^m = \text{sgn}\left[\hat{X}^m W\right] \qquad (1)$$

where sgn (α) = 1 for $\alpha \geq 0$ and sgn (α) = -1 otherwise. Based on column-wise decomposition, the task of designing an AM becomes that of designing P two-category classification weight vectors W_i such that

$$\text{sgn}\left[\hat{X}^m W_i\right] = y_i^m \quad m = 1, 2, \cdots, M \quad i = 1, 2, \cdots, P \qquad (2)$$

That is, $y_i^m X^m W_i > 0$, for m = 1, 2, ..., P. With the normalisation $Z_i^m = y_i^m X^m$, we can look for a weight vector W_i such that $Z_i^m W_i > 0$, m = 1, 2, ..., M, i = 1, 2, ..., P.

Many algorithms can be used to find the parameters of the classifier, such as the perceptron [4], pseudoinverse (PI) [2], least mean square [5], and the Ho-Kashyap (HK) rules [6]. Most of these algorithms are tuned to generate a discriminant surface that is optimal in the mean square error sense, but not tuned to minimise the classification error rate. These methods are therefore not optimal for designing an AM with maximal basin of attraction.

The equation $g(Z_i^m) = Z^m W_i = 0$ defines the hyperplane decision surface that separates the y_i^ms of 1s from y_i^ms of -1s. Instead of merely requiring that the sign of $Z^m W_i$ be positive, we will produce a more reliable AM if we require that the function value $g(Z^m)$ be larger than some margin. The quantity $D(W_i) = \min_m Z^m W_i / \|W_i\|$ can be called *stability*, for it represents the smallest distance of all the Z^m from the plane perpendicular to W_i [3, 7]. Thus $D(W_i)$ will be a measure of how well the hyperplane is selected, for it represents the tolerable noise margin of the AM. A maximal $D(W_i)$ design for each output neuron of an AM would lead to the *optimal stability* solution [7]. Recently, Krauth and Mézard [7] proposed the minimum-overlap (MO) learning rule for finding the

optimal stability design of AMs. This is a perceptron-type algorithm given by

$$\begin{aligned} &\text{maximise} \quad \Delta > 0 \\ &\text{such that} \quad Z^m \cdot W_i - \Delta \geq 0 \quad m = 1, 2, \cdots, M \\ &\qquad \|W_i\| = 1 \end{aligned} \tag{3}$$

The minimal-overlap learning procedure is summarised as follows:

Step 1: Initialise $W_i = 0$ and select a fixed positive number c.

Step 2: Determine a pattern Z^μ that has minimum overlap with W_i:

$$Z^\mu \cdot W_i = \min_m Z^m \cdot W_i \tag{4}$$

Step 3: If $Z^\mu \cdot W_i > c$, then go to Step 4. Otherwise, update W_i by

$$W_i = W_i + \frac{1}{N}(Z^\mu)^T \tag{5}$$

Go to Step 2.

Step 4: Normalise W_i to unity

$$W_i = \frac{W_i}{\|W_i\|} \tag{6}$$

and stop.

The stability determined by this algorithm is

$$D(W_i) = \min_m Z^m \cdot W_i \tag{7}$$

It has been proved [7] that this algorithm will converge in a finite number of time steps and that for $c \to \infty$ the stability will be maximised, i.e.

$$D(W_i) \to \max_{W_i} D(W_i) \quad \text{as} \quad c \to \infty \tag{8}$$

The above learning rule maximises the stability so that the basin of attraction of the AM is as large as possible. However, this design rule leaves the valuable parameter threshold unused, and thus leaves room to further improve the performance of AMs. (The inclusion of the threshold parameter does not introduce any overhead to the neural system, for the threshold is an intrinsic parameter of each neuron.) To this end, this Letter will modify the MO algorithm to include the threshold parameter.

Modified method: Let $X'^m = [X^m, -1]$ and $W_i' = [W_i^T, s_i]^T$, where s_i denotes the threshold. In analogy with eqn. 8, the minimum-overlap rule is employed to find the solution that maximises the value

$$\begin{aligned} \frac{1}{\|W_i'\|} \min_m Z'^m W_i' &= \frac{1}{\|[W_i, s_i]\|} \min_m [Z^m, -y_i^m] \begin{bmatrix} W_i \\ s_i \end{bmatrix} \\ &= \frac{1}{\|[W_i, s_i]\|} \min_m Z^m W_i - y_i^m s_i \end{aligned} \tag{9}$$

This value is not the stability, because the stability considers only the norm of the weight vector, $\|W_i\|$. Rather, eqn. 9 considers both the norm of the weight vector and the threshold, $\|W_i, s_i\|$. To obtain a better assessment of stability, a scheme that will reduce the effect of the threshold in eqn. 9 has to be devised.

We now modify the augmented representation to $X''^m = [X^m, -a]$, where a is a large number, so the augmented weight vector should become $\{W_i''\} = \{W_i, s_i/a\}$. In this way, the inaccuracy of the stability assessment induced by the threshold will be lessened. The minimum-overlap algorithm can then be applied to maximise the quantity

$$\begin{aligned} \frac{1}{\|W_i''\|} \min_m Z''^m W_i'' &= \frac{1}{\|[W_i, \frac{s_i}{a}]\|} \min_m [Z^m, -a y_i^m] \begin{bmatrix} W_i \\ \frac{s_i}{a} \end{bmatrix} \\ &= \frac{1}{\|[W_i, \frac{s_i}{a}]\|} \min_m Z^m W_i - y_i^m s_i \end{aligned} \tag{10}$$

Comparing eqns. 9 and 10, we see that the effect of the threshold in eqn. 10 is less than that in eqn. 9. Moreover the larger the parameter a is, the less significant the effect of the threshold becomes. Ideally, if $a \to \infty$, then

$$\|W_i''\| = \left\| \left[W_i, \frac{s_i}{a} \right] \right\| \to \|W_i\| \tag{11}$$

The quantity in eqn. 10 will approximate the stability, i.e.

$$\frac{1}{\|W_i''\|} \min_m Z''^m W_i'' \to \frac{1}{\|W_i\|} \min_m Z^m W_i - y_i^m s_i \tag{12}$$

as $a \to \infty$. Thus, we can approximate the optimal stability solution by using this modified method. After learning with the minimum-overlap rule, the optimal weight vector W_i^* and threshold s_i^* are given by

$$\begin{aligned} W_i^* &\simeq [w_1, w_2, \cdots, w_n]^T \\ s^* &\simeq a \cdot w_{n+1} \end{aligned} \tag{13}$$

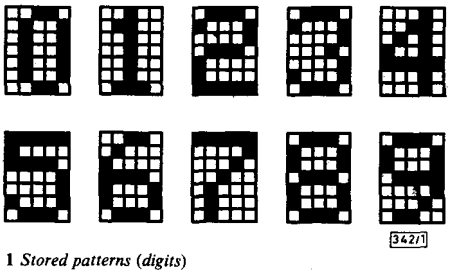

Fig. 1 *Stored patterns (digits)*

Computer simulation: We stored 10 digits [8] with 35 (7 × 5) bipolar neurons, as shown in Fig. 1. In this simulation, the initial weights were all set equal to zero. Parameters are fixed at $a = 20$ and $c = 10^5$. The performance of the recall rate against the Hamming distance is plotted in Fig. 2. The results obtained by the PI, NH, and MO methods are also shown for comparison. Note that the solution found by the NH algorithm (with margin vector set to vector 1) is the same as that by the PI method. For the 35 output neurons, each neuron's stability by the proposed scheme was, as expected, the largest and was, on average, 7 and 10% greater than that by the MO and PI methods, respectively. The modified method thus leads to a significant improvement in the performance of the AM.

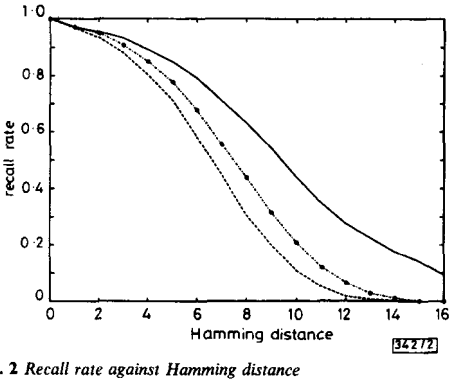

Fig. 2 *Recall rate against Hamming distance*

——— modified method
·-·-O-·-· minimum-overlap
--------- pseudo inverse, HK

In another example, alphabet patterns A–J were stored and the performance showed a similar trend.

Conclusions: To design a Hopfield-type associative memory, we have proposed modifying the minimum-overlap algorithm to include a threshold parameter. A technique that diminishes the effect of the threshold on the minimum-overlap learning algorithm is presented. Using the modified learning rule, we have designed an AM with maximal basin of attraction. The simulation results show that the modified method constructs a larger basin of attraction than do other well-known methods.

31 August 1993
Electronics Letters Online No: 19931389

J.-Y. Chang and C.-C. Wu (*Department of Control Engineering, National Chiao Tung University, Hsinchu, Taiwan 30050, Republic of China*)

References

1 HOPFIELD, J.J.: 'Neurons with graded response have collective computational properties like those two-state neurons', *Proc. Nat. Acad. Sci. USA*, 1984, **81**, pp. 3088–3092
2 KOHONEN, T.: 'Self-organization and associative memory' (Springer-Verlag, New York, 1984)
3 HERTZ, J., KROGH, A., and PALMER, G.: 'Introduction to the theory of neural computation' (Addison-Wesley, Reading, MA, 1991)
4 DUDA, R., and HART, P.: 'Pattern classification and scene analysis' (Wiley-Interscience, New York, 1973)
5 WIDROW, B., and LEHR, M.A.: '30 years of adaptive neural networks: perceptron, madaline, and backpropagation', *Proc. IEEE*, 1990, **78**, (9), pp. 1415–1442
6 HO, Y.-C., and KASHYAP, R.L.: 'An algorithm for linear inequality and its application', *IEEE Trans.*, 1965, **EC-14**, pp. 683–688
7 KRAUTH, W., and MÉZARD, M.: 'Learning algorithm with optimal stability in neural networks', *J. Phys. A: Math. Gen.*, 1987, **20**, pp. L745–L752
8 HASSOUN, M.H., and SONG, J.: 'Adaptive Ho-Kashyap rules for perception training', *IEEE Trans.*, 1992, **NN-3**, (1), pp. 51–61

Enhanced kernel estimation technique for pattern classification

K.P. Lam and E. Horne

Indexing terms:

The Letter reports the application of a nonparametric density estimation technique, the generalised K-nearest-neighbour (K-NN) method, to a novel pattern classifier for binary images. In addition to offering an improved error rate performance over the fixed kernel method previously adopted, the method can be used to measure the inherent difficulty of a pattern classification problem because the nearest-neighbour error rate bounds the Bayes rate.

Introduction: Nonparametric density estimators employing the fixed kernel approach offer simplicity in design and have been shown to provide good results in a unified BWS-FWS pattern classifier [1]. However, the choice of the smoothing parameter, defined as h in the density estimator,

$$\Psi(X) = \frac{1}{Nh^d} \sum_{i=1}^{i=N} K\left(\frac{X - x_i}{h}\right) \qquad (1)$$

[2] has a decisive effect on the performance of such a technique, especially in classification problems where the reduction of error rates relies on an accurate estimation of the density functions $p(X|\omega_j)$ of individual pattern classes. From a statistical standpoint, the 'optimum' h value will always be some kind of a compromise. In regions of high density h will be too large and will produce an 'out-of-focus' estimate with poor resolution, whereas in regions of low density h will be too small, resulting in an estimate which is subjected to excessive statistical variability.

The nearest-neighbour technique [2] represents an attempt to overcome these problems by adapting the amount of smoothing to the 'local' density of data. In essence, given a class and its associated variables, rather than applying the same h everywhere in the feature space of X, the Euclidean distances of some predefined neighbourhood of X within the data samples are adopted as a metric for controlling the amount of smoothing.

A study of the characteristics of K-NN applied to the BWS-FWS classifier architecture is presented in this Letter.

Generalised K-NN density estimator: Given a set of N data samples of x_i in the feature space from which the density function $\psi(V)$, defined above, is to be determined, the K-NN estimator has the general form

$$\Psi(t) = \frac{1}{Nd_{k(t)}} \sum_{i=1}^{i=N} K\left(\frac{t - x_i}{d_k(t)}\right) \qquad (2)$$

where $d_k(t)$ represents the distance from t to its kth nearest neighbour points of the samples and $K(x)$ is a kernel function integrating to unity. To allow useful comparison with previous work [1], the same kernel is chosen and is given here for reference.

$$K(z) = K_C(z)K_D(z)$$

$$K_C(z) = K_D(z) = \begin{cases} 0.5 & \text{if } |z_C| \text{ or } |z_D| > 1.0 \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

It is worth noting that $\psi(t)$ in eqn. 2 is precisely the kernel estimate evaluated at t with window width $d_k(t)$. Clearly, the amount of smoothing at any particular point depends on the density of observations near that point, with the overall effect governed by the choice of the integer k.

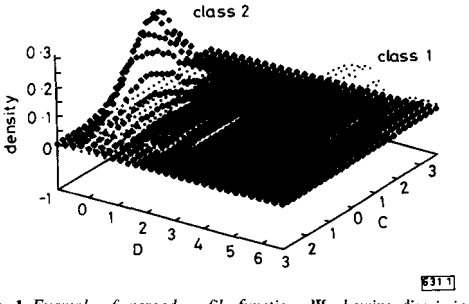

Fig. 1 *Example of merged profile functions Ψ, showing discrimination analysis of two-class case (classes 1 and 2)*

The unknown patterns are taken from class 2
· '1. density'
◊ '2. density'

The example in Fig. 1 demonstrates the generalised K-NN approach for class discrimination.

Implementation and performance evaluation: A 10 way BWS-FWS classifier network was constructed and used for classifying machine-printed postcode numerals [3]. A rotation error estimator [4] was used to obtain the overall performance. The results are presented in conjunction with the error rates obtained from the fixed kernel approach in Table 1.

Table 1: Results in [1] and error rates and values of optimal distance k

Class	Fixed kernel method		GeneralisedK-NNmethod	
	% Error	Value of h_o	% Error	Value of k
'0'	6.33	1.79	4.00	16
'1'	0.00	1.25	0.33	5
'2'	0.33	1.25	0.00	12
'3'	4.67	0.99	7.67	16
'4'	0.00	1.15	0.00	5
'5'	3.34	1.25	4.34	15
'6'	0.33	1.78	0.33	8
'7'	1.34	1.15	1.67	11
'8'	8.47	1.00	5.67	10
'9'	0.33	1.15	0.00	10
	Overallerror rate = 2.514%		Overall error rate = 2.401%	