Adaptive learning based on bit-significance optimization of the Hopfield model and its electro-optical implementation for correlated images

Soo-Young Lee, Chang-Sup Shim, Ju-Seog Jang, and Sang-Yung Shin

Korea Advanced Institute of Science and Technology, Department of Electrical Engineering

P.O. Box 150, Cheongryang, Seoul, Korea

ABSTRACT

Introducing and optimizing bit-significance to the Hopfield model, ten highly correlated binary images, i.e., numbers "0" to "9", are successfully stored and retrieved in a 6x8 node system. Unlike many other neural networks models, this model has stronger error correction capability for correlated images such as "6", "8", "3", and "9". The bit-significance optimization is regarded as an adaptive learning process based on least-mean-square error algorithm, and may be implemented with another neural nets optimizer. A design for electro-optic implementation including the adaptive optimization networks is also introduced.

It is well known that the Hopfield model can successfully retrieve data only when the amount of stored data is much smaller than the number of neurons and when the stored vectors are (pseudo) orthogonal.[1,2] For many practical systems such as number and/or alphabet recognition, the orthogonality is not satisfied, and more complex models have been proposed.[3-5] Athale introduced attention in correlation domain[3], and we had introduced relative significance of each bit, and shown how *a priori* knowledge on bit significance improves error-corrective association performance.[4] In this Letter we show that optimization of the bit significance for given stored images removes or at least relaxes the orthogonality condition and increases storage capacity. This is an adaptive learning process for bit significance based on least-mean-square(LMS) algorithm, and may be implemented with Widrow-Hoff optimizer.

We start with a simple modification on the retrieving algorithm of the Hopfield model to obtain

$$\hat{v}_i^t = \sum_{j=1}^{N} T_{ij} w_j (2v_j^t - 1) \tag{1}$$

where w_j is positive bit significance of the jth bit of image vectors and N is number of neurons. The interconnection matrix T and thresholding operation are same as the Hopfield model but we do not set the diagonal terms zero. If the input vector v^t is one of the stored images, one obtains

$$\hat{v}_i^t = (2v_i^t - 1)\sum_{j=1}^{N} w_j + \sum_{s \neq t} [(2v_i^s - 1)\sum_{j=1}^{N} w_j (2v_j^s - 1)(2v_j^t - 1)] \tag{2}$$

where unipolar binary images are assumed and the second term corresponds to correlation noise. Because one has (N-1) degrees of freedom to select relative significance w_j's, the correlation noise term can be minimized for given stored image vectors.

Following LMS error algorithm one defines error E as

$$E(w) = \frac{1}{2}\sum_{s=1}^{M}\sum_{t=1}^{s-1}[\sum_{j=1}^{N}(2v_j^s-1)(2v_j^t-1)w_j]^2/(\sum_{j=1}^{N}w_j)^2,$$
$$= \frac{1}{2N_w^2}w^TSw, \quad (3)$$

and finds $w = [w_1 w_2 \cdots w_N]^T$ for minimum error. Here the error is summed over all (v^s, v^t) pairs and normalized over $N_w^2 \equiv (\sum w_j)^2$, and S is an NxN positive semi-definite matrix with elements

$$S_{ij} \equiv \sum_{s=1}^{M}\sum_{t=1}^{s-1}(2v_i^s-1)(2v_i^t-1)(2v_j^s-1)(2v_j^t-1)$$
$$= \frac{1}{2}(T_{ij}^2-M). \quad (4)$$

A simple iterative minimization procedure, steepest descent method, updates **w** by $w^{k+1} = w^k - \alpha g^k$, where superscript k denotes k'th iterative solutions and **g** is gradient of error, i.e. $g = \nabla_w E = (Sw - \mathbf{1}EN_w)/N_w^2$ where $\mathbf{1} = [1\ 1\ 1 \cdots 1]^T$. Convergence of the above iteration is guaranteed with proper choice of α.[6] Optimum α_{opt} may be calculated by minimizing $E(w-\alpha g)$ and approximated as $\alpha_{opt} = g^TSw/g^TSg$ for small **w** corrections. The denominator never goes to zero unless $g = 0$, and α_{opt} exists for all non-optimum **w**. Unlike other adaptive neural nets, this model does not need previously learned images to learn a new image. It only needs T and M. This learning property greatly reduces learning time as well as storage requirements.

Increased associative performance of this bit-significance optimization model is demonstrated by computer simulation. Ten highly correlated images are learned recursively from "0" to "9" in 6x8 nodes associative memories. Error correction probabilities versus Hamming distance are plotted for this model and single-layer perceptron in Fig. 1. 1000 input images are randomly generated to satisfy required Hamming distance with each of the stored images, fed to the model, and their convergence characteristics are collected. It is interesting to see that correlated images such as "6", "8", "3", and "9" have much better error-correction performance than uncorrelated images such as "1" and "7" in our model. Perceptron, like many other neural nets models, works poorly for correlated images as shown in Fig. 1(b). This is the most important characteristics of the optimized-bit-significance model, which minimizes sum of errors for all stored image pairs. In case that some stored images are highly correlated, errors contributed from this images should be smaller to minimize overall error. False converging rate is also much smaller for our model than for the perceptron.

Electro-optic implementation has great potential for neural networks.[7] Calculations requiring N^2 order of multiplications for N neuron systems are done by optics, while electronics perform more complicated functions. The optics portion of the design shown in Fig. 2 is quite similar to optical implementation of perceptron in Ref. 8. To take advantage of the optical matrix-vector multiplier special consideration is required for negative numbers. The Hopfield interconnection weights T_{ij}'s are modified to

$$T_{ij}^* = \sum_{s=1}^{M}\frac{1}{2}[(2v_i^s-1)(2v_j^s-1)+1], \quad (5)$$

and implemented by sum of two vector outer products, i.e. $T^* = \sum_{s=1}^{M}[v^s(v^s)^T + \bar{v}^s(\bar{v}^s)^T]$ where $\bar{v}^s = 1-v^s$. Now S_{ij} becomes

$$S_{ij} = 2(T_{ij}^*)^2 - 2MT_{ij}^* + \frac{M(M-1)}{2}. \quad (6)$$

In Fig. 2 all lenses for proper optics are not shown for brevity, and all the SLMs and photodetectors (PDs) are connected to and controlled by a personal computer. The learning and recall procedures are summarized in Table 1. At learning stage SLM1 and SLM2 are first set to a learning vector v^s to create vector outer products on the detector side of the MSLM. Then the two SLMs are switched to v^t. In both cases the MSLM is set to ADD mode operation and generates a new T^*. At other times the MSLM is set to READOUT mode. SLM3, 4, and 5 are set to ON, w, and ON respectively, and PD1 and PD2 are detecting T^*w and S^*w, respectively. Here $S^*_{ij} = (T^*_{ij})^2$. Gradient g and new bit-significance vector w are calculated by the personal computer, and SLM4 is updated to w. For the α_{opt} one resets SLM5 to g and gets $g^T S^* w$ at PD1. Then resets SLM4 to g, and gets $g^T S^* g$ at PD1. Once $S^* w$ is calculated, $w^T S^* w$ is requiring N multiplications only.

For proper recall one also needs to modify Eq. (1) into

$$\hat{v}_i^t = 4\sum_j T^*_{ij} w_j v_j^t - 2\sum_j T^*_{ij} w_j - 2M\sum_j w_j v_j^t + M\sum_j w_j \qquad (7)$$

and make each calculation contain positive values only. At recall stage SLM3 and 4 are set to an input v^t and w, respectively, and PD1 detects the first term in the right hand side of Eq. (7). The 2nd term has been calculated at the learning stage and stored in the personal computer. The other terms may be calculated electronically, and with proper thresholding a new v^t is obtained.

In conclusion we propose bit-significance optimization of the Hopfield model and a design for electro-optic implementation. Computer simulation shows this model works much better than standard Hopfield model and single layer perceptron. The nice feature of this new model includes superior error correction performance for highly correlated images, and fast learning process without priviously-learned images. Extension to multilayer structures is also under investigation.

FIGURE CAPTIONS

Fig.1 Error correction probability vs. Hamming distance for 10 stored images (a) optimized bit-significance model (b) single-layer perceptron.

Fig.2 Design for electro-optic hybrid implementation of the optimized-bit-significance Hopfield model.

REFERENCES

1. D. Psaltis and N. Farhat, Opt. Lett. **10**, 98 (1985).
2. B. Macukow and H. H. Arsenault, Appl. Opt. **26**, 34 (1987).
3. R. A. Athale, H. H. Szu, and C. B. Friedlander, Opt. Lett. **11**, 482 (1986).
4. S.-Y. Lee, J.-S. Jang, S.-Y. Shin, and C.-S. Shim, Appl. Opt. **27**, 1921 (1988).
5. J.-S. Jang, S.-Y. Shin, and S.-Y. Lee, Opt. Lett **13**, 693 (1988).
6. T. Kohonen, *Self-Organization and Associative Memory* pp. 177-182, Springer-Verlag, Berlin (1988).
7. N. H. Farhat, Appl. Opt. **26**, 5093 (1987).
8. A. D. Fisher, W. L. Lippincott, and J. N. Lee, Appl. Opt. **26**, 5039 (1987).

Table 1. Learning and Recall Procedures

	MSLM	SLM1	SLM2	SLM3	SLM4	SLM5	PD1	PD2
Learning								
1. Calculate $v^i(v^i)^t$	ADD	v^i	v^i					
2. Calculate $v^i(v^i)^t$	ADD	v^i	v^i					
3. Calculate T^*w and S^*w	READOUT			ON	w	ON	T^*w	S^*w
4. Calculate w^tS^*w	READOUT			ON	w	w	T^*w	w^tS^*w
5. Calculate g^tS^*w	READOUT			ON	w	g	T^*w	g^tS^*w
6. Calculate T^*g and g^tT^*g	READOUT			ON	g	g	T^*g	g^tT^*g
7. Calculate a new w. If not converged, go to step 3.								
Recall								
1. Calculate $\sum T_{ij} w_j v_j^i$	READOUT			v^i	w		$\sum T_{ij} w_j v_j^i$	
2. Calculate a new v^i. If not converged, go to step 1.