ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 223-228

An Asymmetric Associative Memory Model Based on Relaxation Labeling Processes

Marcello Pelillo and Anna M. Fanelli

Dipartimento di Informatica
Università di Bari
Via E. Orabona 4, 70126 Bari, Italy

Abstract. Relaxation labeling processes are a family of parallel distributed processing models widely popular in computer vision. One of the most remarkable features of these models is that certain local convergence results are known to hold under unrestricted circumstances. In this paper we take advantage of these properties to develop a novel kind of (multivalued) associative memory that employs asymmetric connections. Experiments are presented which confirm the validity of the proposed approach.

1. Introduction

Since the influential work of Hopfield [1], there has been an explosion of interest in the study of neural network models of associative memory. However, despite the manifested inspiration from neuroscience, it is generally agreed that the Hopfield model turns out to be unsatisfactory from a biological standpoint, especially because of the (essential) requirement that neurons be connected in a symmetrical fashion.

In the computer vision domain, another class of parallel distributed models has long been successfully employed in numerous applications [2]. These algorithms, known as *relaxation labeling processes*, were developed to solve certain problems arising in the 3D interpretation of ambiguous line drawings [3], and many authors have recently stressed that intriguing similarities exist between them and both artificial and biological neural systems [4], [5]. This claim is further supported by recent work showing that relaxation labeling processes not only possess interesting learning abilities [6], but are also capable of solving hard optimization problems [7]. The most notable feature of relaxation labeling, which makes it especially related with the Hopfield neural model, is that when symmetric connections are used there exists an energy function that is minimized as the process evolves [8]. Interestingly, the same kind of dynamical behavior is also exhibited by the original relaxation scheme which was developed in a heuristic fashion [9]. However, some intriguing differences exist as well. In particular, a general local convergence result has been proven to hold even when the symmetry condition is relaxed [8], [10].

Based on such properties of relaxation labeling processes, in this paper a novel kind of associative memory model is developed. The problem of loading patterns in the proposed memory turns out to be equivalent to imposing a certain consistency condition and this, in turn, amounts to solving a (sparse)

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 223-228

system of linear inequalities. One of the most interesting features of the relaxation labeling memory is that no constraint is imposed on the structure of the connection matrix and, in fact, the algorithm *does* work with asymmetric connections. Moreover, owing to the particular representation of memory vectors employed, the approach is completely general and allows us to develop multivalued memories in a very natural way. Some experimental results are reported which demonstrate the validity of the proposed model.

2. Relaxation Labeling Processes and Their Properties

Relaxation labeling processes were developed to solve the so-called (continuous) labeling problem, where one has to assign labels to objects so as to satisfy a set of domain-specific constraints. Let $\boldsymbol{B} = \{b_1, \cdots, b_n\}$ and $\boldsymbol{\Lambda} = \{0, \cdots, m-1\}$ denote respectively the set of objects and the set of labels for the problem at hand. Moreover, let the constraints be expressed in terms of a four-dimensional matrix of real-valued compatibility coefficients R: the element $r_{ij}(\lambda, \mu)$ measures the strength of compatibility between the two hypotheses "λ is on object b_i" and "μ is on object b_j." High values mean compatibility while low values mean incompatibility.

Let $p_i(\lambda)$ represent the degree of confidence of the hypothesis "label λ is on object b_i." It is assumed that $p_i(\lambda) \geq 0$ and $\Sigma_\lambda p_i(\lambda) = 1$, so that the vector $\overline{p}_i = (p_i(0), \cdots, p_i(m-1))^T$ can be considered as the probability distribution of labels for b_i. By putting together the $\overline{p}_i$'s we obtain a weighted labeling assignment for the objects of $\boldsymbol{B}$ that will be denoted by $\overline{p}$, and will be conveniently considered as an $n \times m$ matrix. We find it useful to introduce the space of weighted labeling assignments:

$$\mathbb{K} = \left\{ \overline{p} \in \mathbb{R}^{nm} : p_i(\lambda) \geq 0, \text{ all } i, \lambda \text{ and } \sum_\lambda p_i(\lambda) = 1, \text{ all } i \right\}$$

which is a linear convex set of $\mathbb{R}^{nm}$. Every vertex of $\mathbb{K}$ represents an *unambiguous* labeling assignment which assigns exactly one label to each object. The set of these labelings is denoted by $\mathbb{K}^* = \{\overline{p} \in \mathbb{K} : p_i(\lambda) = 0 \text{ or } 1\}$.

Hummel and Zucker [8] developed a general theory of consistency for the labeling problem which is the basis of the work reported here. The entire development of the theory is a generalization of the notion of consistency for unambiguous labelings, which is more easily understood. Consider a labeling $\overline{p} \in \mathbb{K}$. The degree of agreement between the hypothesis that b_i is labeled with λ and the context can be quantified by a linear *support* function

$$q_i(\lambda; \overline{p}) = \sum_j \sum_\mu r_{ij}(\lambda, \mu) p_j(\mu) \qquad \text{all } i, \lambda \,. \tag{1}$$

Now, let $\overline{p} \in \mathbb{K}^*$ be an unambiguous labeling, and let $\lambda(i)$ denote the label assigned to b_i by $\overline{p}$ (i.e. $p_i(\lambda(i)) = 1$). It seems reasonable to say that $\overline{p}$ is consistent if and only if the assigned label of each object receives the greatest support at that object. This corresponds to having

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 223-228

$$q_i(\lambda;\overline{p}) \leq q_i(\lambda(i);\overline{p}) \ , \qquad \text{all } i,\lambda \tag{2}$$

or equivalently $\Sigma_\lambda v_i(\lambda)q_i(\lambda;\overline{p}) \leq \Sigma_\lambda p_i(\lambda)q_i(\lambda;\overline{p})$, for all $\overline{v} \in \mathbb{K}^*$. By analogy, a labeling $\overline{p}$ is said to be *consistent* provided that

$$\sum_\lambda v_i(\lambda)q_i(\lambda;\overline{p}) \leq \sum_\lambda p_i(\lambda)q_i(\lambda;\overline{p}) \ , \qquad \text{all } i \tag{3}$$

for all $\overline{v} \in \mathbb{K}$. Furthermore, if the inequalities in (3) are strict, for all $\overline{v} \neq \overline{p}$, then $\overline{p}$ is said to be *strictly consistent.* It can be easily shown that for unambiguous labelings the conditions (2) and (3) are equivalent. After defining the notion of consistency, Hummel and Zucker [8] showed that when the compatibility matrix R happens to be symmetric, then a sufficient condition for a labeling $\overline{p}$ to be consistent is that it is a local minimum of the following "energy" function, which is a measure of labeling's (in)consistency

$$A(\overline{p}) = -\sum_{i,j}\sum_{\lambda,\mu} r_{ij}(\lambda,\mu)p_i(\lambda)p_j(\mu) \ . \tag{4}$$

A relaxation labeling process takes as input an initial labeling $\overline{p}^{(0)} \in \mathbb{K}$ and iteratively adjusts it taking into account the compatibility model. The most popular relaxation scheme, which was used in the experiments reported here, is

$$p_i^{(t+1)}(\lambda) = p_i^{(t)}(\lambda)q_i^{(t)}(\lambda) \Big/ \sum_\mu p_i^{(t)}(\mu)q_i^{(t)}(\mu) \qquad \text{all } i,\lambda \tag{5}$$

provided that the compatibilities are nonnegative. This corresponds to the original nonlinear formulas developed heuristically by Rosenfeld *et al.* [3].

Recently, it has been shown that, despite its completely heuristic derivation, the original relaxation scheme (5) possesses in fact a number of interesting properties. First, when R is symmetric then A turns out to be a Liapunov function for the process which means that it is monotonically decreasing along nonconstant trajectories [9]. Second, and even more interestingly, it can be proven that strictly consistent labelings act as local attractors for the process defined in (5) whether or not the matrix R happens to be symmetric [10, Theorem 10]. These properties are the basis for the associative memory model developed in this paper.

3. The Memory Model

In this section it is shown how to build up an associative memory that is based upon relaxation labeling processes. In our discussion it will be assumed that an m-ary ($m \geq 2$) memory is to be constructed. Let the patterns to be stored have the form $\xi = \xi_1\xi_2\cdots\xi_n$, where $\xi_i \in \{0,\cdots,m-1\}$ for all $i = 1...n$. Note that there exists a one-to-one correspondence between ξ and the following unambiguous labeling $\overline{p}(\xi) \in \mathbb{K}^* \subset \mathbb{R}^{nm}$:

$$p_i(\lambda;\xi) = \begin{cases} 1, & \text{if } \lambda = \xi_i \\ 0, & \text{if } \lambda \neq \xi_i \end{cases} \ , \qquad \text{all } i,\lambda \ .$$

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 223-228

Based on this representation, the proposed associative memory model consists of an $n \times m$ densely interconnected relaxation labeling network, the connection strengths between units being determined by the compatibility matrix R. The unit indexed (i, λ) updates its state according to formulas (1) and (5), and its activation value $p_i(\lambda)$ can therefore be thought of as the "probability" that λ be the correct value for word's site i. The network is started with an initial configuration $\overline{p}^{(0)} \in \mathbb{K}$, and evolves until a stable state is reached. From the preceding section we know that if $\overline{p}^{(0)}$ is sufficiently close to a strictly consistent labeling (which must be an unambiguous one) then the network will eventually approach it, whatever the structure of the connection matrix R is. If the resulting unambiguous labeling corresponds to a memory pattern, we say that the network has recalled that pattern. This suggests that, to store a vector ξ in the network, we need to find a compatibility matrix R that makes the corresponding labeling $\overline{p}(\xi)$ strictly consistent (we shall find it helpful to "linearize" the matrix R and consider it as an n^2m^2-dimensional vector $\overline{r}$).

To make the discussion more formal, suppose that P memory vectors $\xi^{(1)}, \cdots, \xi^{(P)}$ are to be stored, and consider the corresponding unambiguous labelings $\overline{p}(\xi^{(1)}), \cdots, \overline{p}(\xi^{(P)})$ as defined previously. For the labelings $\overline{p}(\xi^{(\gamma)})$, $\gamma = 1 \ldots P$, to be strictly consistent the following relation must be true

$$\sum_j r_{ij}(\lambda, \xi_j^{(\gamma)}) - \sum_j r_{ij}(\xi_i^{(\gamma)}, \xi_j^{(\gamma)}) < 0 \qquad (6)$$

for all $\lambda \neq \xi_i^{(\gamma)}$, $i = 1 \ldots n$, and $\gamma = 1 \ldots P$. This is a system of $Pn(m-1)$ linear inequalities in the n^2m^2 unknowns $\{r_{ij}(\lambda, \mu)\}$, which can be compactly represented as $C\overline{r} < \overline{0}$, where C is the $Pn(m-1) \times n^2m^2$ matrix defined as

$$C(\gamma, i, \lambda; j, k, \mu, \eta) = \begin{cases} +1, & \text{if } (j = i) \wedge (\mu = \lambda) \wedge (\eta = \xi_k^{(\gamma)}) \\ -1, & \text{if } (j = i) \wedge (\mu = \xi_j^{(\gamma)}) \wedge (\eta = \xi_k^{(\gamma)}) \\ 0, & \text{otherwise} \end{cases}$$

(for convenience we use a three-component index for the rows and a four-component index for the columns), $\overline{r}$ is the unknown compatibility vector, and $\overline{0}$ is the null vector. In practice, it is customary when solving such systems to introduce a "margin" [11], which also ensures larger basins of attraction [12]. Accordingly, our system is rewritten as $C\overline{r} \leq \kappa\overline{1} < \overline{0}$, where κ is some predetermined constant, and $\overline{1}$ is the unity vector.

One interesting algorithm for solving systems of linear inequalities was proposed by Eremin [13]. Its main attractive feature, which distinguishes it from standard methods [11], is that it *does* solve the system when it happens to be compatible and automatically yields the "best" approximation solution (in the sense of Chebyshev) when this is not the case. More specifically, the algorithm iteratively computes a solution of $C\overline{r} - \kappa\overline{1} \leq \epsilon_0\overline{1}$, where ϵ_0 is the *defect* of the original system, i.e. the smallest ϵ for which the system $C\overline{r} - \kappa\overline{1} \leq \epsilon\overline{1}$ is compatible. (Note that this corresponds to solving the original system when it has a solution.) Space limitations do not allow us to provide a description of Eremin's method which is, however, readily available [13].

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 223-228

4. Simulation Results

To assess the validity of our model, some experiments were performed aimed at testing its storing and error-correction capabilities. The learning set used in the study was taken from [14], and consisted of ten binary digits on a 5×5 matrix (Fig. 1). The training was carried out according to the Eremin procedure, using a margin $\kappa = -10$, which appeared to be near-optimal.

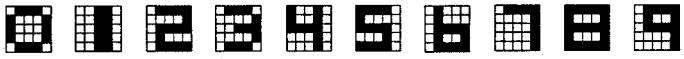

Fig. 1. Training set used in the experiments.

For each memory pattern, ten perturbed versions were generated by randomly flipping exactly d bits (for $d = 0, 1, \ldots$) and the corresponding unambiguous labelings were obtained. Also, since unambiguous labelings turn out to be fixed points for the relaxation scheme (5), a further Gaussian noise with mean 0 and variance 0.1 was inflicted (this was followed by a successive normalization step to ensure that the noisy labelings still belonged to $\mathbb{K}$). The network was then allowed to iterate for (at most) 1,000 steps, and the final (weighted) labelings were converted into binary patterns by a simple maxima selection rule. The resulting patterns were then compared with the original ones, and a *success* was recorded when a perfect match was achieved. Fig. 2. shows the percentage of successful recalls, as a function of d.

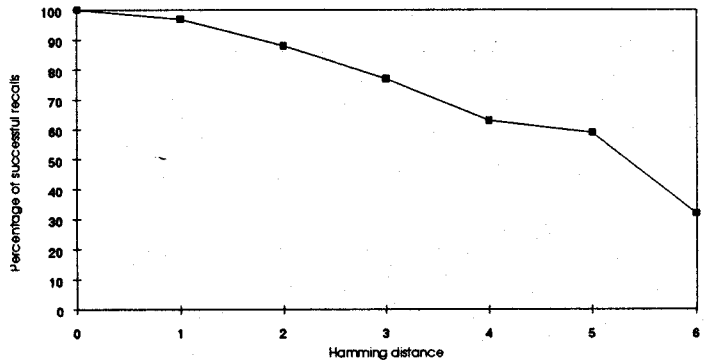

Fig. 2. Error-correction performance as a function of the Hamming distance.

As can be seen, the network exhibits good error-correction capabilities, even for (moderately) high values of d. The results are far superior to those obtained with standard Hebb-trained Hopfield networks and compare favorably with the ones reported in [14, Fig. 4], using a Hopfield network trained with an optimal perceptron-based learning procedure.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 223-228

5. Conclusions

In this paper, a novel approach to building associative memories has been presented which is based upon certain dynamical properties of relaxation labeling processes. The approach is quite general and allows us to naturally develop multivalued memories, which are believed to be more closely related to biology and turn out to be useful in practical applications. The storing process is formulated in terms of a system of linear inequalities that we solve by means of an efficient relaxation-style algorithm. Preliminary results over a set of binary patterns have been presented which demonstrate the effectiveness of the proposed approach. Further work is in progress aimed at estimating the storage capacity of the model and evaluating the number of spurious states.

References

1. J.J. Hopfield (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **79**, 95-110.
2. J. Kittler and J. Illingworth (1985). Relaxation labeling algorithms – A review. *Image Vision Comput.* **3**, 206-216.
3. A. Rosenfeld, R.A. Hummel, and S.W. Zucker (1976). Scene labeling by relaxation operations. *IEEE Trans. Syst. Man Cybern.* **6**(6), 420-433.
4. S.S. Yu and W.H. Tsai (1992). Relaxation by the Hopfield neural network. *Pattern Recognition* **25**, 197-209.
5. S.W. Zucker, A. Dobbins, and L. Iverson (1989). Two stages of curve detection suggest two styles of visual computation. *Neural Comp.* **1**, 68-81.
6. M. Pelillo and M. Refice (1994). Learning compatibility coefficients for relaxation labeling processes. *IEEE Trans. Pattern Anal. Machine Intell.* **16**(9), 933-945.
7. M. Pelillo (1993). Relaxation labeling processes for the traveling salesman problem. *Proc. IJCNN-93*, Nagoya, Japan, 2429-2432.
8. R.A. Hummel and S.W. Zucker (1983). On the foundations of relaxation labeling processes. *IEEE Trans. Pattern Anal. Machine Intell.* **5**(3), 267-287.
9. M. Pelillo (1994). Nonlinear relaxation labeling as growth transformation. *Proc. 12th ICPR*, Jerusalem, Israel, II:201-206.
10. T. Elfving and J.-O. Eklundh (1982). Some properties of stochastic labeling procedures. *Comput. Graph. Image Processing* **20**, 158-170.
11. R.O. Duda and P.E. Hart (1973). *Pattern Classification and Scene Analysis.* Wiley, New York.
12. B.M. Forrest (1988). Content addressability and learning in neural networks. *J. Phys. A: Math. Gen.* **21**, 245-255.
13. I.I. Eremin (1962). Iteration method for Chebyshev approximations for sets of incompatible linear inequalities. *Soviet Math. Doklady* **3**, 570-572.
14. X. Zhuang, Y. Huang, and F.A. Yu (1994). Design of Hopfield content-addressable memories. *IEEE Trans. Signal Process.* **42**(2), 492-495.