PERGAMON

Neural Networks 12 (1999) 1207–1212

Neural Networks

www.elsevier.com/locate/neunet

Contributed article

The asymptotic memory capacity of the generalized Hopfield network

Jinwen Ma*

Institute of Mathematics, Shantou University, Shantou, Guangdoug 515063, People's Republic of China

Received 8 August 1997; received in revised form 21 May 1999; accepted 21 May 1999

Abstract

This paper presents a theoretical analysis on the asymptotic memory capacity of the generalized Hopfield network. The perceptron learning scheme is proposed to store sample patterns as the stable states in a generalized Hopfield network. We have obtained that $(n-1)$ and $2n$ are a lower and an upper bound of the asymptotic memory capacity of the network of n neurons, respectively, which shows that the generalized Hopfield network can store the larger number of sample patterns than Hopfield network.

Keywords: Neural network; Hopfield network; Asymptotic memory capacity; Associative memory; Perceptron learning algorithm

1. Introduction

When Hopfield network was proposed as an associative memory model in 1982, the sum-of-outer product scheme was applied to store the sample patterns (Hopfield, 1982). Hopfield demonstrated by computer simulation that the network with n neurons could store about $0.15n$ patterns in the form of the stable states. It is now well known that the asymptotic memory capacity of Hopfield network with n neurons is $n/(4 \log n)$ patterns (McMliece et al., 1987).

Hopfield network is a single layer recurrent network of n bipolar (or binary) neurons uniquely defined by $(\mathbf{W}, \theta)$ where $\mathbf{W}$ is a symmetric zero-diagonal real weight matrix, and θ is a real threshold vector. If the weight matrix is changed to be an asymmetric and zero-diagonal one, the network is usually called an asymmetric Hopfield network. In this paper, we define a generalized Hopfield network (GHN) to be such kind of a network with a general (asymmetric or symmetric) and zero-diagonal real weight matrix.

Recent researches (Ma, 1997) show that the GHN having stable states can be stable in the same way as a Hopfield network. Thus it is possible to apply this neural architecture to associative memory with some learning scheme which enables a set of prescribed patterns as the stable states of a GHN. Moreover, several such learning schemes on the GHNs for associative memory have already been established (see, e.g. Gardner, 1988; Wang et al., 1993). However, the memory capacity of the GHN with any learning scheme has not been investigated in depth. From the literature of neural networks, the following theoretical results are related to the memory capacity of the GHN.

Abu-Mostafa and Jacques (1985) defined the memory capacity as the maximal number of arbitrary state patterns that can be stable in a GHN of n neurons and proved[1] that it is bounded by n. In fact, this deterministic definition of memory capacity is too strict since we can easily verify that any pair of the two state patterns with one Hamming distance cannot be stable in any GHN. Therefore the memory capacity defined by this deterministic formulation is insignificant and the obtained bound is loose and useless.

The other way to define the memory capacity of some kind of neural network (with some learning scheme) is via the probability sequence $P(m,n)$ that m random state patterns can be stable in a choice of the neural network of n neurons (by the learning scheme). Venkatesh and Psaltis (1989) defined a function $C(n)$ as the (asymptotic) memory capacity if, and only if, for every $\lambda \in (0,1)$, as $n \to \infty$, $P(m,n)$ approaches one whenever $m \le (1-\lambda)C(n)$, and zero whenever $m \ge (1+\lambda)C(n)$. By this definition, they found that $C(n) = 2n$ is the asymptotic memory capacity of the recurrent network defined by a general weight matrix and a threshold vector (Venkatesh, 1987; Venkatesh and Psaltis, 1989). In a special case that the threshold vector is zero, it was proved that $C(n) = n$ under each of the spectral strategies (Venkatesh & Psaltis, 1989). Obviously, the

* Tel.: + 86-754-2902170, 2903473; fax: + 86-754-2510654, 2510517.

[1] Since the weight matrix of the network in the theorem of Abu-Mostafa and Jacques (1985) is just a real-valued zero-diagonal matrix, the model of the network is actually a GHN.

PII: S0893-6080(99)00042-8

recurrent network is also a generalization of Hopfield network, but the diagonal elements of the weight matrix are not necessarily zero. Thus it is different from the GHN defined in this paper.

Here we prefer the model of the GHN that remains $w_{ii} = 0$ for the two reasons: (1) It is proved that the GHN with nonnegative weights is stable in randomly asynchronous mode and it is also shown by simulation experiments that almost any GHN having stable states is stable in randomly asynchronous mode (Ma, 1997). By these results, we can consider that the GHN maintains some important properties of the stability of Hopfield network for associative memory. (2) When w_{ii} is restricted to be zero, the network is easy to be implemented for the applications.

However, the restriction that $w_{ii} = 0$, actually brings the difficulty on solving the asymptotic memory capacity and we cannot use the results obtained by Venkatesh and Psaltis (1989). In this paper we will use a method of combinatorial analysis to study the asymptotic memory capacity of the GHN.

The main contribution of this paper is obtaining lower and upper bounds of the asymptotic memory capacity of the GHN. In Section 2, we will propose the main theorem after a brief description of the GHN and the perceptron learning scheme. The proof of the main theorem is given in Section 3. A brief conclusion is given in Section 4.

2. The main theorem

We first give the mathematical model of a GHN. A GHN is composed of n interconnected neurons with $(\mathbf{W}, \theta)$ where $\mathbf{W}$ is an $n \times n$ zero-diagonal matrix with element $w_{i,j}$ denoting the weight on the connection from neuron j to neuron i; and θ is a vector of dimension n with component θ_i denoting the threshold of neuron i. For simplicity, we let $\theta_i = 0$, $i = 1, 2, \ldots, n$ in this paper.

Every neuron can be in one of two possible states, either 1 or -1. The state of neuron i at time t is denoted by $x_i(t)$. The state of the network at time t is denoted by the vector $X(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^{\mathrm{T}}$. The state of ʝat time of storage of the GHN as follows:

$$P(m,n) = P(\{\{X^1, X^2, \ldots, X^m\} : \{X^1, X^2, \ldots, X^m\} \text{ is storable}\}),$$

where $m, n \in \quad = \{1, 2, \ldots\}$. Obviously, $P(m,n)$ decreases with m.

We now introduce the mathematical definition of the asymptotic memory capacity of the GHN based on $P(m,n)$ as follows:

Definition 2. An integer function $C(n)$ is the asymptotic memory capacity of the GHN if the following conditions hold:
(1).

$$\lim_{n\to\infty} P(m,n) = 1 \tag{4}$$

whenever $m \le C(n)$;
(2).

$$\lim_{n\to\infty} \inf P(m,n) < 1 \tag{5}$$

whenever $m > C(n)$.

We say $C(n)$ is a lower bound of the asymptotic memory capacity of the GHN if it satisfies the first condition; and that $C(n)$ is an upper bound of the asymptotic memory capacity if Eq. (5) holds whenever $m \ge C(n)$.

It is clear that this probabilistic definition of the asymptotic memory capacity is different from the definition of Venkatesh and Psaltis (1989) (see in Section 1). We propose this probabilistic definition in order to overcome the two weaknesses of Venkatesh and Psaltis' definition. First, it cannot be proved that the existence of the asymptotic memory capacity of the GHN by Venkatesh and Psaltis' definition. Second, even if there exists the asymptotic memory capacity in this case, it may be not unique under Venkatesh and Psaltis' definition. However, under this probabilistic definition and by the decrease of $P(m,n)$ with m, we can easily prove that there exists a unique asymptotic memory capacity (function) of the GHN. Further, this probabilistic definition of the asymptotic memory capacity is consistent with the general understanding. Suppose that $C(n)$ is the asymptotic memory capacity function of the GHN. If $m \le C(n)$, almost all choices of m patterns can be made stable in a GHN. Otherwise if $m > C(n)$, the number of the choices of m patterns that cannot be made stable in any GHN is of order

$$\binom{2^n}{m}$$

i.e. the number of all choices of m patterns with a finite number of n. Then $C(n)$ is the maximal number of random state patterns that can be stable in a GHN with probability one. Therefore it is consistent with the general understanding and this probabilistic definition of the asymptotic memory capacity (function) is more reasonable and applicable.

We now propose our main theorem as follows:

Theorem 1. *Supposing $P(m,n)$ is the probability sequence of storage of the GHN, we have*
(i).

$$\lim_{n\to\infty} P(n-1, n) = 1;$$

(ii).

$$\lim_{n\to\infty} \inf P(2n, n) \le \frac{1}{2}.$$

The proof of the main theorem is given in next section. We now discuss the significance of the theorem. Because $P(m,n)$ decreases with m, we have by the main theorem that $(n-1)$ and $2n$ are a lower and an upper bound of the asymptotic memory capacity function of the GHN, respectively. Since there exists the asymptotic memory capacity of the GHN $C(n)$, then $C(n) \ge n-1$. Therefore $C(n)$ is much greater than $n/(4\log n)$—the asymptotic memory capacity of Hopfield network of n neurons with the sum-of-outer product scheme. On the contrary, $C(n) < 2n$, which seems reasonable since each neuron of the network can store at most $2n$ patterns when the threshold value is not necessarily zero.

3. The proof of the main theorem

In this section, we will prove the main theorem. The basic difficulty to prove the theorem comes from the fact that W must have zero diagonal. When w_{ii} is not necessarily zero, the proof of relation (i) is closely related to the question of computing the *Vapnik dimension* of the linear classifier (Pollard, 1989). In this case, things are not difficult (linear separability of $n-1$ bipolar vectors in $\mathbb{R}^n$). But when $w_{ii} = 0$, we arrive at the problem of checking the linear separability of a dichotomy of $n-1$ vectors in $\mathbb{R}^{n-1}$, which causes the difficulty, and the vectors are no longer in general position. In order to overcome the difficulty, we will use a sufficient condition of storage to estimate the number of the storable sample sets. By the combinatorial analyses of the number of the storable sample sets over the total number of the sample sets, we will complete the proof of relation (i). On the contrary, we will use Cover's inequality (Cover, 1965) to prove relation (ii).

We first give some lemmas and begin with a sufficient condition of storage under the perceptron learning scheme.

Lemma 1. *Consider a sample set $\{X^1, X^2, \ldots, X^{n-1}\}$. If the same component of each of the vectors $X^1, X^2, \ldots, X^{n-1}$ is* *deleted and the $n-1$ remaining* [redacted]*)-dim vectors are linearly independent then* $\{X^1, X^2,$ [redacted]$^{-1}\}$ *is stor*[redacted]

[redacted]**roof.** Let

[redacted]$^i = [x_{i,1}, x_{i,2}, \ldots, x_{i,n}]^{\mathrm{T}}$ [redacted] $n-1).$ (6)

[redacted] As for $\{X^1, X^2$ [redacted]$\}$ and [redacted] ith neuron of the [redacted]etwork $\mathbf{N} =$ [redacted]$\mathbf{O})$, we constr[redacted]e following system [redacted]f linear [redacted]ions:

[redacted] $w_{i,1} + \cdots + x_{1,i-1}w_{i,i-1} + x_{1,i+1}w_{i,i+1} + \cdots + x_{1,n}w_{i,n} = x_{1,i}$

$$x_{2,1}w_{i,1} + \cdots + x_{2,i-1}w_{i,i-1} + x_{2,i+1}w_{i,i+1} + \cdots + x_{2,n}w_{i,n} = x_{2,i}$$

$$\cdots$$

$$x_{n-1,1}\dot{w}$$

where $w_{i,1}, \ldots, w_{i,i-1}, w_{i,i+1}, \ldots, w_{i,n}$ are $(n-1)$ unknown numbers.

Using the condition of the lemma, it is deduced that the rank of the system matrix of linear equations Eq. (7) is $(n-1)$. Thus the linear equations have a unique solution of $w_{i,1}, \ldots, w_{i,i-1}, w_{i,i+1}, \ldots, w_{i,n}$. In this way for all the neurons, we can obtain $\mathbf{W}$. According to Eqs. (2) and (7), $X^1, X^2, \ldots, X^{n-1}$ are the stable states of the obtained network $\mathbf{N} = (\mathbf{W}, \mathbf{O})$. Therefore $\{X^1, X^2, \ldots, X^{n-1}\}$ is storable and the proof is completed. □

Lemma 2. *Let*

$$A_n = |\{(a_{i,j})_{n\times n} : a_{i,j} \in \{1,-1\} \text{ and } \mathrm{rank}(a_{i,j})_{n\times n} \le n-1\}|,$$

where $|B|$ is the number of the elements of the set B. Then we have

$$\lim_{n\to\infty} \left(\frac{A_n}{2^{n^2}}\right) = 0. \quad (8)$$

The proof is analogous to that of Komlos' theorem (Komlos, 1967) (Komlos' result is for $n \times n$ $\{0,1\}$ matrices, but it holds well for $n \times n$ $\{-1,1\}$ matrices).

For any m-set of points (vectors) $\subset \mathbb{R}^n$, let $\mathscr{B}(\ )$ denote the family of dichotomies of that are homogeneous linearly separable. Here a dichotomy $(\ ^+, \ ^-)$ belongs to $\mathscr{B}(\ )$ if, and only if, there exists a weight vector $W = [w_1, w_2, \ldots, w_n]^{\mathrm{T}} \in \mathbb{R}^n$ such that $(X = [x_1, \ldots, x_n]^{\mathrm{T}} \in \mathbb{R}^n)$

$$\mathrm{sgn}(W^{\mathrm{T}}X) = \mathrm{sgn}\left(\sum_{i=1}^{n} w_i x_i\right) = \begin{cases} 1 & \text{if } X \in \ ^+; \\ -1 & \text{if } X \in \ ^-. \end{cases} \quad (9)$$

The following estimate for the number of homogeneous linearly separable dichotomies was given and proved by Cover (1965).

Lemma 3. *Let* $\subset \mathbb{R}^n$ *be an m-set of points. The following estimate holds*:

$$B(m,n) \text{ [redacted] } \le C(m,n) = 2\sum_{i=0}^{n-1} C_i^{m-1}. \quad (10)$$

Now we [redacted]dy to prove the main theorem.

The Proof of the Main Theorem. (i) Assuming that $X^1, X^2, \ldots, X^{n-1}$ are represented as Eq. (3), we introduce the following symbols.

$$\mathscr{D}_{(n-1)} = \{\{X^1, \ldots, X^{n-1}\} : \{X^1, \ldots, X^{n-1}\} \text{ is not storable}\},$$

$\mathscr{D}^*_{(n-1)}$ is the complementary set of $\mathscr{D}_{(n-1)}$, and

$$D_{(n-1)} = |\mathscr{D}_{(n-1)}|; \qquad D^*_{(n-1)} = |\mathscr{D}^*_{(n-1)}|.$$

$$E_{(n-1)\times n} = \begin{pmatrix} e_{1,1} & e_{1,2} & \cdots & e_{1,n} \\ e_{2,1} & e_{2,2} & \cdots & e_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n-1,1} & e_{n-1,2} & \cdots & e_{n-1,n} \end{pmatrix}$$

$$E_{(n-1)\times n}(i) = \begin{pmatrix} e_{1,1} & \cdots & e_{1,i-1} & e_{1,i+1} & \cdots & e_{1,n} \\ e_{2,1} & \cdots & e_{2,i-1} & e_{2,i+1} & \cdots & e_{2,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ e_{n-1,1} & \cdots & e_{n-1,i-1} & e_{n-1,i+1} & \cdots & e_{n-1,n} \end{pmatrix}$$

where $e_{i,j} \in \{1,-1\}, E_{(n-1)\times n}(i)$ is defined for $i = 1, 2, \ldots, n.$

$$\mathscr{E}^*_{(n-1)} = \{E_{(n-1)\times n} : \mathrm{rank}\ E_{(n-1)\times n}(i) = n-1, i = 1, 2, \ldots, n\},$$

$\mathscr{E}_{(n-1)}$ is the complementary set of $\mathscr{E}^*_{(n-1)}$, and

$$E_{(n-1)} = |\mathscr{E}_{(n-1)}|; \qquad E^*_{(n-1)} = |\mathscr{E}^*_{(n-1)}|.$$

We consider the following matrix constructed by $X^1, X^2, \ldots, X^{n-1}$ as

$$X_{(n-1)\times n} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-1,1} & x_{n-1,2} & \cdots & x_{n-1,n} \end{pmatrix}.$$

Then it is an $E_{(n-1)\times n}$ matrix. According to Lemma 1, we have the following inequality:

$$(n-1)! \times D_{(n-1)} \leq E_{(n-1)}. \tag{11}$$

We now estimate $E_{(n-1)}$ in the following. Let

$$A_{n,k} = \left|\{(a_{i,j})_{n\times n} : a_{i,j} \in \{1,-1\} \text{ and rank } (a_{i,j})_{n\times n} = k\}\right|,$$

where $k = 1, 2, \ldots, n$;

$$A_n = \sum_{k=1}^{n-1} A_{n,k}$$

and $E_{(n-1)\times(n-1)} = E_{(n-1)\times n}(n)$. Then the number of elements of $\mathscr{E}_{(n-1)}$ can be estimated from $E_{(n-1)\times(n-1)}$ by extending a column vector as follows:

(1) If rank $E_{(n-1)\times(n-1)} < n-1$, obviously, $E_{(n-1)\times n} \in \mathscr{E}_{(n-1)}$. Define $E^1_{(n-1)}$ as the number of the matrices of $E_{(n-1)\times n}$ in that case. Then

$$E^1_{(n-1)} = A_{n-1,1}2^{n-1} + A_{n-1,2}2^{n-1} + \cdots + A_{n-1,n-2}2^{n-1}$$
$$= A_{n-1}2^{n-1}.$$

(2) If rank $E_{(n-1)\times(n-1)} = n-1$ and $E_{(n-1)\times n} \in \mathscr{E}_{(n-1)}$, then the nth column vector of $E_{(n-1)\times n}$ must be able to be represented as a linear combination of the $(n-1)$ column vectors of $E_{(n-1)\times(n-1)}$. Define $E^2_{(n-1)}$ as the number of the matrices of $E_{(n-1)\times n}$ in this case. Let $e_1, e_2, \ldots, e_{n-1}$ be the column vectors of $E_{(n-1)\times(n-1)}$, then $\{e_1, e_2, \ldots, e_{n-1}\}$ is a base of $\mathbb{R}^{n-1}$. Thus any $U \in \mathbb{R}^{n-1}$ has a unique representation of linear combination on the base $\{e_1, e_2, \ldots, e_{n-1}\}$, that is

$$U = \sum_{j=1}^{n-1} \alpha_j e_j. \tag{12}$$

U is called a non-zero linear combination of $\{e_1, e_2, \ldots, e_{n-1}\}$ if $\alpha_j \neq 0$ for each $j = 1, 2, \ldots, n-1$; and we let $\mathscr{F}(e_1, e_2, \ldots, e_{n-1}) = \{U \in \{1,-1\}^{n-1} : U$ is a non-zero linear combination of $\{e_1, e_2, \ldots, e_{n-1}\}\}$. We define for $e_i \in \{1,-1\}^{n-1}, i = 1, \ldots, k$ $\Theta_k = \{(e_1, e_2, \ldots, e_k) : e_1, e_2, \ldots, e_k$ are linear independent in $\mathbb{R}^{(n-1)}\}$; and Θ_k^* is the complementary set of Θ_k. Now we have

$$E^2_{(n-1)} = \sum_{(e_1,\ldots,e_{n-1})\in\Theta_{n-1}} \left|\{U \in \{1,-1\}^{n-1} : U \notin \mathscr{F}(e_1, \ldots, e_{n-1})\}\right|.$$

In order to estimate $E^2_{(n-1)}$, we further define for $k \leq n-1$

$$\bar{\ }(e_1, e_2, \ldots, e_k) = \ (e_1, e_2, \ldots, e_k) \cap \{1,-1\}^{(n-1)},$$

where $\ (e_1, e_2, \ldots, e_k)$ is the linear spanning space of $\{e_1, e_2, \ldots, e_k\}$.

Then we have

$$E^2_{(n-1)} = \sum_{(e_1,\ldots,e_{n-1})\in\Theta_{n-1}} \left|\{U \in \{1,-1\}^{n-1} : U \notin \mathscr{F}(e_1, \ldots, e_{n-1})\}\right|$$
$$= \sum_{(e_1,\ldots,e_{n-2})\in\Theta_{n-2}} \sum_{e_{n-1}\notin \bar{\ }(e_1,\ldots,e_{n-2})} \left|\{U : U \in \bar{\ }(e_1, \ldots, e_{n-2})\}\right| = \sum_{(e_1,\ldots,e_{n-2},U)\in\Theta^*_{n-1}} \left|\{e_{n-1} \in \{1,-1\}^{n-1} : e_{n-1} \notin \bar{\ }(e_1, \ldots, e_{n-2})\}\right|$$
$$\leq 2^{n-1}A_{n-1,n-2}.$$

Summing up the results of the two cases, we have

$$E_{(n-1)} = E^1_{(n-1)} + E^2_{(n-1)} \leq 2^{n-1}A_{n-1} + 2^{n-1}A_{n-1,n-2}$$
$$\leq 2^n A_{n-1}.$$

According to Lemma 2 and the fact

$$\lim_{n\to\infty}\left(\frac{2^{n^2}}{P^n_{2^n}}\right) = 1 \qquad (P^m_n = n(n-1)\cdots(n-m+1)); \tag{13}$$

we have

$$1 - P(n-1,n) = \frac{D_{n-1}}{\binom{2^n}{n-1}} = \frac{(n-1)!D_{n-1}}{(n-1)!\binom{2^n}{n-1}}$$
$$= \frac{(n-1)!D_{n-1}}{P^{n-1}_{2^n}} = \left(\frac{2^{n(n-1)}}{P^{n-1}_{2^n}}\right) \times \left(\frac{(n-1)!D_{n-1}}{2^{n(n-1)}}\right)$$
$$\leq \left(\frac{2^{n^2}}{P^n_{2^n}}\right) \times \left(\frac{E_{n-1}}{2^{n(n-1)}}\right) \leq 2\left(\frac{2^{n^2}}{P^n_{2^n}}\right) \times \left(\frac{A_{n-1}}{2^{(n-1)^2}}\right)$$
$$\longrightarrow 0 (n \longrightarrow \infty).$$

Therefore we have

$$\lim_{n\to\infty} P(n-1,n) = 1.$$

(ii) For a sample set $\{X^1, X^2, \ldots, X^{2n}\}$, we let

$X^i(1$

matrix

$$X_{2n\times n} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{2n,1} & x_{2n,2} & \cdots & x_{2n,n} \end{pmatrix} = \begin{pmatrix} x_{1,1} & X^1(1) \\ x_{2,1} & X^2(1) \\ \vdots & \vdots \\ x_{2n,1} & X^{2n}(1) \end{pmatrix}.$$

According to Lemma 3 and the fact $C(2n, n-1) < 2^{2n-1}$, we have

$$\begin{aligned}(2n)!D^*_{2n} &\leq 2^{2n\times(2n-1)}C(2n, n-1) \leq 2^{2n\times(2n-1)}2^{2n-1} \\ &= 2^{(2n)^2-1}\end{aligned}$$

and

$$P(2n, n) = \frac{D^*_{2n}}{\binom{2^n}{2n}} =$$