On Recognition Capability of Hopfield Networks[1]

V. N. Koshelev and B. V. Kryzhanovskii

Scientific Council on Cybernetic, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 117968 Russia
Institute of Optical-Neural Technologies, Russian Academy of Sciences, ul. Vavilova 44/2, Moscow, 117333 Russia
e-mail: mail@valery.msk.ru; kryzhanov@mtu-net.ru

Abstract—The Hopfield associative memory (HAM) is a network with data storage capability. The mathematical apparatus of the HAM is based on linear matrices, which makes the HAM and attractive tool in many applications. In this paper, the concept of HAM is considered as a possible approach to learning pattern recognition. A standard randomization technique is developed to estimate the information capacity and identification efficiency of the HAM. An upper bound for the probability of identification error is found as a function of pattern clusters and the probability of distortion within a cluster.

The term "formal neural network" conventionally refers to a structure comprising N node-neurons joined by edge-synapses [1, 4, 6]. The nth neuron is treated as a binary element taking the values

$$x_n \in \{-1, +1\}, \quad n \in \overline{1, N} \tag{1}$$

and the ijth synapse, as a conductor connecting the ith neuron to the jth neuron and having the conductivity

$$c_{ij} \in R, \quad (i, j) \in \overline{1, N} \times \overline{1, N} \tag{2}$$

Thus, a neural network is determined by the pair

$$(N, C) \tag{3}$$

where N is the size of a network and

$$C \in [c_{ij}] \tag{4}$$

is the matrix of its conductivities.

The network operates at two moments of time, initial and terminal. At the initial moment, the network input, i.e., the initial numerical values of its neurons, is determined by a binary sequence received from the outside and called an input excitation. Under the synchronous action of the network elements, the state of the network changes from initial to terminal; the terminal state also has the form of binary sequence. Based on these definitions, a neural-network concept of recognition is constructed, according to which the network must assign the same output excitation to all "similar" input excitations; i.e., it must associate similar input excitations to the same template. In the simplest form, this concept assumes the presence of the "environment" where M phenomena can occur. The random occurrence of the mth phenomenon excites the initial state

$$x_m^N \theta^N = [x_{m1}\theta_1, x_{m2}\theta_2, \ldots, x_{mN}\theta_N] \tag{5}$$

where

$$x_m^N \in \{-1, +1\}^N, \quad m \in \overline{1, M} \tag{6}$$

is a subset of vertices of the N-dimensional unit cube (these vertices play the role of templates) and

$$\theta^N \in \{-1, +1\}^N \tag{7}$$

is "multiplicative noise" consisting of N independent equidistributed random variables

$$\theta_n = \begin{cases} +1, 1-p \\ -1, p \end{cases} \quad \forall n \in \overline{1, N} \tag{8}$$

For instance, $m \in \overline{1, M}$ may be the number of a species of animals (say, cats); then, x_m^N is a standard set of features of cats as a species, and $x_m^N \theta^N$ is the set of features of an individual, i.e., a particular cat taken at random. Let the $M \times N$-sizes matrix

$$B = \begin{bmatrix} x_1^N \\ \vdots \\ x_M^N \end{bmatrix} \tag{9}$$

be the set of M templates. We say that the quintuple

$$(N, M, C, B, p) \tag{10}$$

specifies a neural-network recognizing device in the form of the chain of mappings

$$\begin{aligned} m &\longrightarrow x_m^N \longrightarrow x_m^N\theta^N \longrightarrow (x_m^N\theta^N)\mathbf{C} \longrightarrow \\ &\longrightarrow \operatorname{sgn}[(x_m^N\theta^N)\mathbf{C}] \equiv x^N \end{aligned} \tag{11}$$

The first three links in this chain, "phenomenon $\longrightarrow$ template $\longrightarrow$ input excitation," are determined by the

[1] This work was partially supported by the Russian Foundation for Basic Research, project no. 00-01-00213.

Received July 30, 1999

properties of the environment, the fourth link is the linear transformation

$$y^N = (x_m^N \theta^N)\mathbf{C} \in R^N \tag{12}$$

of the input excitation (this transformation is determined by the conductivity of network (9)), and the fifth link is the terminal state

$$\operatorname{sgn} y^N = x^N \in \{-1, +1\}^N \tag{13}$$

obtained by applying the componentwise operation

$$\operatorname{sgn} y_n = \begin{cases} +1, & y_n > 0 \\ -1, & y_n \le 0 \end{cases} \tag{14}$$

On the whole, chain of transformations (11) defines a nonlinear operator mapping the space of states $\{-1, +1\}^N$ of the network into itself. We consider this chain as a neural-network recognizing device specified by quintuple (10). This chain is assigned the probability of recognition error

$$\begin{aligned} P_{\text{er}}(N) &= P_{\text{er}}(N; M; C; B; p) \\ &= \sum_{m=1}^{M} M^{-1} \sum_{t^N} p(t^N)\chi\{x^N(\mathbf{C}) \ne x_m^N\} \end{aligned} \tag{15}$$

where

$$x^N(\mathbf{C}) = (x_m^N t^N)\mathbf{C}, \quad p(t^N) = (1-p)^{N^+} p^{N^-} \tag{16}$$

Here, N^+ is the number of +1 in the sequence t^N and $N^- \hat{=} N - N^+$.

Treating quintuple (10) as a function of N, we introduce the notion of effectiveness of scheme (11).

Definition 1. We say that the neural-network recognizing device determined by quintuple (10) is effective if

$$\lim_{N \to \infty} \inf P_{\text{er}}(N) = 0 \tag{17}$$

Thus, the effectiveness of a neural network is understood as the capability to identify observations with given templates. In this sense, the probability of correct identification is a measure of the ability of the network to make associative decisions.

The conjecture that neural networks can manifest elements of associative behavior [4] is confirmed by the example of Hopfield networks [2], which have the special feature: their conductivity matrices are determined by the template matrices according to

$$\mathbf{C} = \mathbf{B}'\mathbf{B} - \mathbf{I} \tag{18}$$

where $\mathbf{I}$ is the identity $N \times N$ matrix and the prime denotes transposition.

Definition 2. Suppose that a template matrix of form (9) is given. A neural network of form (3) whose conductivity matrix (4) is determined by the template matrix according to (18) is called a Hopfield network.

Since the matrix of conductivities is fixed, a Hopfield network is determined by the quadruple

$$(N, M, \mathbf{B}, p) \tag{19}$$

Recognizing device (11) based on a Hopfield network coincides formally with the scheme for a noise-resistant transmission of M messages through a binary channel without memory with multiplicative noise (7). In this scheme, template matrix (9) plays the role of a code book of size M, the third link in (11) corresponds to the output of the transmission channel, and the fourth and fifth links corespond to the two-stage coding by the scheme

$$\{-1, +1\}^N \longrightarrow R^N \longrightarrow \{-1, +1\}^N \tag{20}$$

Accordingly, the probability of recognition error is understood as the probability of decoding error. To estimate the latter, it is natural to apply the method of averaging over the ensemble of code books, i.e., the ensemble of template matrices (9). The consistency of such an approach to neural networks is shown in [3, 5]. In this work, we use the method of averaging over ensemble based on the Chebyshev-Chernov method for estimating large deviation probabilities, which makes the proof rigorous and indicates the possibility of its further generalization.

Let

$$\mathfrak{R} = \begin{bmatrix} \beta_1^N \\ \vdots \\ \beta_M^N \end{bmatrix}, \quad \beta \in \{-1, +1\}^N \tag{21}$$

be a random uniformly distributed $M \times N$ matrix, and let

$$\begin{aligned} &\overline{P_{\text{er}}(N; M; \mathfrak{R}; p)} \\ &= \sum_{m=1}^{M} M^{-1} \sum_{t^N} p(t^n)\chi\{x^N(\mathfrak{R}^N) = \beta^N\} \end{aligned} \tag{22}$$

be the mean probability of recognition error in the Hopfield network generated by random matrix (21).

The main result of this work is the following theorem.

Theorem (on the effectiveness of the randomized Hopfield network). *For arbitrary integer $N \ge 1$ and $M \le 1$ and any p such that $0 \le p \le 1/2$, the mean recognition error over the ensemble of templates in the Hopfield network satisfies the inequality*

$$\overline{P_{\text{er}}(N+1; M+1; \mathfrak{R}; p)} \le N e^{-\frac{N}{2M}(1-2p)^2} \tag{23}$$

To simplify the expression, we have omitted the factor [1 + o(1)] (o(1) $\longrightarrow$ 0 as $M \longrightarrow \infty$ [sign: infinity]) in the exponent. The results of a numerical experiment show that, at $N \geq 100$, the strict inequality sign can be used in (23) with a high accuracy (the error is less than 1%).

REFERENCES

1. McCullach, W.S. and Pitts, W., A Logical Calculus of the Ideas Miniment in Nervous Activity, *Bull. Math. Biophys.*, 1943, vol. 5, pp. 115–133.
2. Hopfield, J.J., Neural Networks and Physical Systems with Emergent Collective Computational Abilities, *Proc. Nat. Acad. Sci. U.S.A.*, 1982, vol. 79, pp. 2554–2558.
3. McEliece, R.J., Posner, E.C., Rodemich, E.R., and Venkatsch, S.S., The Capacity of the Hopfield Associative Memory, *IEEE Trans. Inf. Theory*, 1987, vol. IT-33, pp. 461–482.
4. Hebb, D.O., *The Organization of Behavior,* New York: Wiley, 1949.
5. Kuh, A. and Dickinson, B.W., Information Capacity of Associative Memories, *IEEE Trans. Inf. Theory,* 1989, vol. IT-35, pp. 55–68.
6. Kilkarni, S.L., Lugosi, G., and Venkatesh, S.S., Learning Pattern Classification-A Survey, *IEEE Trans. Inf. Theory,* 1998, vol. 44, pp. 2178–2204.

SPELL: equidistributed, componentwise, corespond, Miniment