Fractal Neural Network

---Computational Performance as an Associative Memory---

Eung-Soo Kim, Masaki Sano and Yasuji Sawada

Research Institute of Electrical Communication
Tohoku University, Sendai 980

(Received January 25, 1993)

Associative memory with fractal connections was studied. We show numerically that the fractal neural network has a higher capability than randomly connected neural network for fractal pattern and fractally localized patterns, but not for random patterns. We discuss this result from a theoretical standpoint.

Recent studies of artificial neural networks have been partly motivated by their resemblance to biological neural systems. However, relatively little knowledge has been obtained concerning network structure and coding principles in biological neural systems. Therefore most studies on associative memory, for instance, concern fully or randomly connected networks as their structures. It is known that the human brain has $10^{10}$ neurons, and each neuron has $10^3$ to $10^4$ connections with other neurons. Even if localized brain function is considered, neither fully connected nor randomly connected structures are biologically plausible. Neuro-physiological studies have revealed recently that the branching pattern of dendrites has fractal properties in various parts of the mammalian brain. On the other hand, it is known that the neural firing rate of an individual neuron is very low compared with its saturation level. This is reminiscent of the characteristic of sparse coding: the total contribution of the excited components is very small compared with the resting components.

Motivated by these observations, we have studied the performance of a recurrent network model of associative memory with fractal connections. However, our goal is not to compose a biological neural network model with a fractal structure but to evaluate the computational performance of the fractally configured network model. For a fractally connected network with $N$ neurons, the ratio of connections for each neuron to the total number of neurons is very small, approaching 0 for a large $N$ limit. The fractal neural network model as an associative memory was studied by Baram, who suggested a structure consisting of small subnetworks interconnected in a layered hierarchy. Merrill and Port proposed a fractal neural network and applied it to a simple generalization problem.

In this paper, we compare the computational performance of a fractal neural network (FNN) with a randomly connected neural network (RNN), using the same number of internal connections. We use random patterns, random fractal patterns and fractally localized patterns for encoding. Since fractal patterns have a very low activity level, they need a sparse coding scheme to be stored. We show that...
FNN has a higher capability than RNN for fractal pattern and fractally localized patterns, but not for random patterns. We discuss this result from a theoretical standpoint.

Random fractal connection was constructed for a $N$ neuron system as follows. We define the spatial coordinate of $i$-th neuron by the $d$-dimensional position vector $r^i = (r^i_1, r^i_2, \ldots, r^i_d)$. If the distance between the $i$-th and $j$-th neurons is $r_{ij} = \|r_i - r_j\| = (\sum_{k=1}^{d}(r^k_i - r^k_j)^2)^{1/2}$, the connection $c_{ij}$ between two neurons is determined as 1, with probability $p(r_{ij})$, and as 0 with probability $1-p(r_{ij})$, where $p(r)$ is defined by a power law distribution function:

$$p(r) = a r^{D_N - d} \quad (r > 0 \text{ and } 0 \leq D_N \leq d)$$ (1)

The connectivity $c_{ij}$ is calculated for all $i$ and $j$. Therefore the number of connections $C(R)$ contained in a $d$-dimensional sphere with radius $R$ is

$$C(R) = \int_0^R p(r) d^d r \sim R^{D_N}$$ (2)

implying that the fractal dimension of synaptic connection distribution is $D_N$, where $d^d r = dr^1 dr^2 \ldots dr^d$. Note that when $D_N = d$ and $a = 1$, $p(r) = 1$ holds, which implies the fully connected condition. Numerical simulations of the $N$ neuron system were performed on a lattice for the condition $d = 1$, using the Euclidean distance $r_{ij} = |r_i - r_j|$. For simplicity, the neuron lattice is defined by the interval $[0, N-1]$, i.e., $r_i = (i-1)$, for $(i=1, \ldots, N)$. For RNN, synaptic connections are symmetric, i.e., $c_{ij} = c_{ji}$.

The model consists of $N$ neurons, each capable of being in one of two states, represented by the values 1 (firing state) and -1 (resting state). We represent the state of the $i$-th neuron as $x_i (i=1, \ldots, N)$. The state of the network is defined by the vector $x = (x_1, \ldots, x_N)^T$, where $T$ denotes transpose. The network operates as an associative memory when the strength of each synapse is determined by $M$, given embedded patterns according to the outer product rule, i.e., Hebbian learning. Here, we write the $i$-th component of the $\mu$-th embedded pattern as $s^\mu_i = \pm 1$ $(i=1, \ldots, N, \mu =1, \ldots, M)$. The average firing rate $\bar{f}$ of embedded patterns is defined by

$$\bar{f} = \frac{1}{NM} \sum_{\mu=1}^{M} \sum_{i=1}^{N} (s^\mu_i + 1)/2.$$ (3)

The random pattern has the firing rate $\bar{f} = 1/2$, while the firing rate for a fractal pattern depends on its fractal dimension $D_p$, because the relations $N \sim R^d$ and \{the number of fired sites\} $\sim R^{D_p}$ hold:

$$\bar{f} \sim \frac{R^{D_p}}{N} \sim \frac{R^{D_p}}{R^d} \sim N^{D_p/d-1}.$$ (4)

Therefore, storing fractal patterns results in sparse coding since in the limit $N \to \infty$, $\bar{f} \to 0$ holds. For learning, we use covariant rule:

$$J_{ij} = \frac{C_{ij}}{N} \sum_{\mu=1}^{M} (s^\mu_i - \bar{s}_i) (s^\mu_j - \bar{s}_j),$$ (5)
where
\[ s_i = \frac{1}{M} \sum_{\mu=1}^{M} s_i^\mu. \] (6)

Then, the fractal patterns are stable states of the network, provided that the threshold is appropriately chosen, for
\[ \theta_i = 4 \bar{f}(1-\bar{f}) s_i. \] (7)

The dynamics of the network, which is appropriate both for dense and sparse codings, is defined by
\[ x'_i = \text{sgn}\left[ \sum_{\mu} J_{ij}(x_j - \bar{s}_j) + \theta_i \right], \] (8)

where \( x \) denotes the present state, and \( x' \) the next state. Note that, in the case \( \bar{f} = 1/2 \), \( \bar{s}_i = \theta_i = 0 \) holds, implying Hebbian learning and standard dynamics. Also in the limiting case of sparse coding, \( \bar{f} \to 0 \), the relation \( \bar{s}_i \to -1 \) and \( \theta \to 0 \) holds, Eq. (8) results in a similar form with generalized Willshaw dynamics,\(^{2,4}\) with 1 and 0 signal value coding instead of 1 and -1.

For test patterns, random patterns, random fractal patterns (random Cantor sets) and fractally localized patterns were used to encode the synaptic strength for each network. The characteristics of random Cantor set and fractally localized test patterns requires some explanation.

One of the simplest and well-known fractals is the Cantor set, which is a finite size fractal consisting of disconnected parts embedded into one-dimensional space \((d=1)\). The Cantor set is constructed based on the subsequent division of intervals generated on the unit interval \([0, 1]\). First \([0, 1]\) is replaced by \(m\) intervals of length \(1/n\) \((m < n)\). Next, this rule is applied to the \(m\) newly created intervals, and the procedure is repeated. As a result we obtain a deterministic fractal whose fractal dimension is \(D_p = \ln m/\ln n\). In general, Cantor sets with various \(m\) and \(n\) can be constructed. For example, keeping \(n=10\) and changing \(m\) between 1 and 9, nine different fractal dimensions \(0 \leq D_p \leq 1\) can be produced. By choosing \(m\) intervals out of \(n\) at random, we constructed various Cantor sets with the same fractal dimension. Calculated random Cantor sets were expanded on the lattice sites by multiplying \([0, 1]\) by \((N-1)\).

Let us consider one-dimensional fractally localized patterns. The fractally localized pattern with fractal dimension \(D_p\) is constructed based on the power law distribution function:
\[ q(r_{ij}) = b r_{ij}^{D_p - d}, \quad (r > 0 \text{ and } 0 \leq D_p \leq d) \] (9)

where \(d=1\) and \(r_{ij} = |r_i - r_j|\), and \(b\) is fixed at 1/2. For the \(\mu\)-th set of pattern \(s^\mu\), we chose \(r_i\) at random and set \(s_i^\mu = 1\), then the \(j\)-th component \(s_j^\mu\) was determined as 1 with probability \(q(r_{ij})\) and as -1 with probability \(1 - q(r_{ij})\).

To compare the computational capability of FNN and RNN with the same number of internal connections, numerical simulations were carried out with \(N=500\), for the range, \(0 < D_p < 1\), of the network dimension, starting from a randomly selected initial state \(s^\mu(\neq s^\mu)\), with parallel dynamics and a finite number of test patterns as
Fig. 1. Memory capacity of FNN and RNN for fractally localized patterns with $D_p=0.4$ and 0.6 pattern dimension, respectively. (a) FNN with $D_p=0.6$, (b) RNN with $D_p=0.6$, (c) FNN with $D_p=0.4$, (d) RNN with $D_p=0.4$. Each line represents an approximation of the capacity for FNN and RNN.

Random Cantor sets as test patterns were generated randomly, subject to the constraint $D_p=\ln 4/\ln 10$. Fractally localized patterns were generated subject to the constraint $D_p=0.4$ and 0.6. The total activity of the fractal patterns depends on their fractal dimension. The number of active components in a Cantor sets and fractally localized patterns becomes negligibly small compared to $N$, for $D_p=\ln 4/\ln 10$ and $D_p=0.4$, respectively. The rate of average activity is 4% for fractally localized patterns of $D_p=0.4$, and 10% for $D_p=0.6$. Even if the networks have the same network dimension, the capacity of the network depends on the pattern dimension $D_p$. Figure 1 shows that the capacity of the network is bigger when $D_p$ is greater, for the same network dimension, $D_N$. Here, we note that even if the Cantor sets and fractally localized patterns have the same pattern dimension, $D_p$, the capacities of the network are different because the firing rates are different. The capacity of FNN and RNN for stability of the memory states is shown in Fig. 1, for fractally localized patterns, and in Fig. 2 for random Cantor sets. We can observe that there is a critical value of the network dimension for memory capacity. The rates of total connectivity of the network for $D_N=0.8, 0.9$ and 0.98 are about 41%, 63% and 91%, respectively. Figures 1 and 2 show that the memory capacity does not increase at all with the increase of the network dimension, until $D_N$ reaches 0.8. When $D_N \geq 0.8$, the memory capacity increases gradually at first, then rapidly when $D_N$ approaches 1 (full connection). These simulation results show that the performance of the FNN is greater than the RNN for fractally represented patterns (random Cantor sets) and fractally localized patterns. We show the memory capacity determined by random patterns.
Fractal Neural Network

Connectivity of network

Fig. 2. Memory capacity of FNN and RNN for random Cantor sets with \( D_p = \ln 4 / \ln 10 \) pattern dimension. (a) □: FNN (solid line), (b)△: RNN (dotted line). Each line represents an approximation of the capacity for FNN and RNN.

Fig. 3. Memory capacity of FNN and RNN for random patterns with \( p = 1/2 \). (a) □: FNN (solid line), (b)△: RNN (dotted line). Each line represents an approximation of the capacity for FNN and RNN.

for FNN and RNN in Fig. 3. The capacities of FNN and RNN are small because there are few effective connections when the network has a low value of dimension \( D_N \). Capacity gradually increases on increasing the network dimension. For
instance, when \( D_N = 0.8 \), the capacity is \( \bar{p} (= M/N) = 0.086 \) for RNN, and 0.080 for FNN. This result shows that the performance of the RNN is greater than the FNN for the random patterns.

Since the number of excited components of fractal patterns is very small compared with the number of neurons \( N \), it is difficult to confirm the results for the asymptotic limits \((N \rightarrow \infty)\) by numerical simulations. However, it is possible to see the effects of fractally configured connections on the capability of an associative memory.

It is known that the memory capacity for random patterns is determined by the average connectivity and is higher for a larger connectivity in case of RNN, where embedded patterns are uniformly encoded in all the synaptic connections.\(^9\) However, the capacity of FNN for correlated patterns such as fractal patterns, where synaptic strengths are not uniformly encoded in space, has not been studied. Since the fractal pattern, for instance, has a stronger correlation for a shorter distance, a connection between two neurons over a short distance has a higher strength than that between two over a long distance. Therefore the effective connectivity which is created by the correlation between synaptic configurations and embedded fractal patterns is different when correlated patterns are stored. Let us consider \( i, j \) pair where \( J_{ij} \neq 0 \). If \( c_{ij} \neq 0 \), the connection between \( i \)-th and \( j \)-th neuron is effective for the function of network, while it is not if \( c_{ij} = 0 \). Suppose \( c_{ij} = 0 \) for all \( i, j \), the capacity is naturally zero. While if \( c_{ij} = 1 \) for all \( i, j \), the network is fully connected and has highest memory capacity. Therefore the memory capacity is considered to be greater if the total effectiveness of the network is higher. The effective connectivity is defined as the expectation of \( J_{ij} \), which is the correlation of the connection and synaptic weight:

\[
\langle J_{ij}\rangle = \frac{1}{N} \sum_{\mu=1}^{M} S_{\mu} S_{\mu} \delta(r - r_{i}) \delta(r'' - r_{j}) d^{d}r d^{d}r'' , \tag{10}
\]

where \( S_{\mu} = s_{\mu} - \bar{s}_{\mu} \), and \( V \) is the volume of the network space. Because \( s_{\mu} = s_{\mu} - \bar{s}_{\mu} \) is either 0 or 1 for the sparse limit, \( \langle J_{ij}\rangle \) is roughly proportional to the number of synaptic connections which are not 0 after learning. Therefore, it is assumed that the higher effective connectivity produces the higher capacity.

**THEOREM**

*If the correlation function \( g(r) \) of embedded patterns is a monotonically decreasing function of the distance \( r \), any FNN has higher effective connectivity \( \langle J_{ij}\rangle \) than a RNN with the same average connectivity, when a covariant learning rule is used.*

Both the correlation function of the embedded patterns, \( g \), and network connectivity, \( c_{ij} \), can be written as functions of distance \( r = \| r_{i} - r_{j} \| \). Therefore, the effective connectivity, Eq. (10), can be evaluated by changing integration variables:

\[
\int_{r} \int_{r} d^{d}r d^{d}r'' = \int_{V} \int_{V} d^{d}r d^{d}r'' , \tag{11}
\]

where \( r = r' - r'' \), \( R \equiv (r' + r'')/2 \). The correlation function of the patterns, \( g \), is given by
Fractal Neural Network

\[ \sum_{\mu=1}^{M} S_{i,\mu}^{s} = \langle S_{i,\mu}^{s} S_{i,\mu}^{s} \rangle = \langle S_{i,\mu}^{s} S_{i,\mu}^{s} \delta (r - \|r_i - r_j\|) \rangle = g(r). \]  

(12)

For FNN, the expectation \( c_{i}^{e} \) is also a function of the distance \( r \):

\[ \langle c_{i}^{e} \delta (r - \|r_i - r_j\|) \rangle = \frac{1}{V} \int_{V} c_{i}^{e} \delta (r - \|r_i - r_j\|) d^{d}r = \rho (r). \]  

(13)

For RNN, \( c_{i}^{e} \) and \( \sum_{\mu=1}^{M} S_{i,\mu}^{s} \) are statistically independent. Thus we obtain

\[ \langle c_{i}^{e} \sum_{\mu=1}^{M} S_{i,\mu}^{s} S_{i,\mu}^{s} \rangle = \langle c_{i}^{e} \rangle \langle \sum_{\mu=1}^{M} S_{i,\mu}^{s} S_{i,\mu}^{s} \rangle = \bar{p} \langle \sum_{\mu=1}^{M} S_{i,\mu}^{s} S_{i,\mu}^{s} \rangle, \]  

(14)

because the total connections are the same for RNN and FNN, i.e., \( \bar{p} = \langle c_{i}^{e} \rangle = \langle c_{i}^{e} \rangle = 1/V \int_{V} \rho (r) d^{d}r \). Evaluating the following equation:

\[ \langle J_{i}^{e} \rangle - \langle J_{i}^{e} \rangle = \langle (c_{i}^{e} - \bar{p}) \sum_{\mu=1}^{M} S_{i,\mu}^{s} S_{i,\mu}^{s} \rangle = \int_{V} \int_{V} (c_{i}^{e} - \bar{p}) g(r) d^{d}r d^{d}r' \]

\[ = \int_{V} d^{d}R \int_{V} (p(r) - \bar{p}) g(r) d^{d}r. \]  

(15)

Equation (15) is positive definite by the following argument.

Let us denote the distance \( r=R_{1} \), such that \( \rho (R_{1}) = \bar{p} \), and \( C = g(R_{1}) \). By subtracting the identity \( \int_{V} (p(r) - \bar{p}) C d^{d}r = 0 \) from both sides of Eq. (15), we obtain

\[ \int_{V} (p(r) - \bar{p}) g(r) d^{d}r = \int_{0}^{R_{1}} (p(r) - \bar{p}) (g(r) - C) d^{d}r \]

\[ + \int_{R_{1}}^{R_{\text{max}}} (p(r) - \bar{p}) (g(r) - C) d^{d}r > 0, \]  

(16)

since \( \{ p(r) - \bar{p} > 0 \} \) and \( \{ g(r) - C > 0 \} \) for \( \{ 0 < r < R_{1} \} \), \( \{ p(r) - \bar{p} < 0 \} \) and \( \{ g(r) - C < 0 \} \) for \( \{ R_{1} < r < R_{\text{max}} \} \) hold. This argument is valid also for any monotonically decreasing function \( \rho (r) \). The equality holds only if \( g(r) = C \) or \( \rho (r) = \bar{p} \).

For arbitrary patterns whose auto-correlation function is a monotonically decreasing function of distance \( r \), including fractal patterns, a FNN of arbitrary fractal dimension has higher effective connectivity than a RNN. However this argument does not explain why a RNN has higher capacity than a FNN for random patterns, for which \( g(r) = 0 \) holds implying \( \langle J_{i}^{e} \rangle = \langle J_{i}^{e} \rangle \). This may come from the fact that \( \langle J_{i} \rangle \) does not hold the meaning of effective connectivity any more for dense coding, since \( S_{i} = s_{i} - \bar{s}_{i} \) is either \(-1\) or \(1\). This discrepancy may also come from multi-step correlation in the dynamics, or the asymmetrical connections in FNN instead of symmetrical connections as in RNN.

References