Multiplicative Neural Network with Thresholds

Leonid Litinskii and Magomed Malsagov
Center for Optical Neural Technologies,
Scientific Research Institute for System Analysis of Russian Academy of Sciences,
Vavilov Str, Moscow, Russia

Keywords: Hopfield-type Network, Multiplicative Connections, Ground State.

Abstract: The memory of Hopfield-type neural nets is understood as the ground state of the net – a set of configurations providing a global energy minimum. The use of thresholds allows good control over the ground state. It is possible to build multiplicative networks with the degeneracy of the ground state exceeding considerably the dimensionality of the problem (that is, the net memory can be much greater than the dimensionality of the problem). The paper considers the potentials and limitations of the approach.

1 INTRODUCTION

Let us consider a Hopfield-type neural network with a multiplicative connection matrix $M_{ij} = (1-\delta_{ij})u_i u_j$, $i, j = 1,.., p$. Here δ_{ij} is the Kronecker symbol, p is the space dimensionality, real numbers u_i are the coordinates of normalized vector $\mathbf{u} = (u_1,...,u_p)$: $\|\mathbf{u}\|^2 = p$. The fixed point of the net is a configuration whose binary coordinates are the signs of coordinates of vector $\mathbf{u}$:

$$\mathbf{s} = (s_1, s_2, ..., s_p) \text{ is a fixed point;}$$
$$s_i \equiv \mathrm{sgn}(u_i),\ i = 1,.., p.$$

The set of fixed points changes significantly if we define the dynamics of the same matrix by using thresholds T_i, which are not only non-zero, but also proportional to coordinates u_i :

$$\begin{cases} J_{ij} = f(x)M_{ij}, \quad T_i = g(x)u_i, \\ s_i(\tau+1) = \mathrm{sgn}\left(\sum_{j=1}^{p} J_{ij}s_j(\tau) + T_i\right) \end{cases} \tag{1}$$

Being functions of parameter x, multipliers $f(x)$ and $g(x)$ themselves serve as free parameters of the model. $s_i(\tau)$ is the i-th coordinate of configuration $\mathbf{s}(\tau)$ determining the state of the net at time τ. The arrangement of the set of fixed points of this sort of net is more complicated and more interesting. It turns out to be possible to determine fully the configuration sets that bring the energy functional to a global minimum. Such configurations are usually called the ground state of the net (the term is borrowed from physics). It is the ground state that is regarded as the memory of a net: it is also the case with the Hebb matrix and projection connection matrix (Hertz et al., 1991).

Given model (1), it is possible to determine analytically the dependence of the ground state on external parameters f, g, x and $\mathbf{u}$. It is possible to control the ground state by varying external parameters. In short, the findings are as follows. Generally speaking, the whole set of 2^p configurations $\mathbf{s}$ falls into sets of configurations that are equally distant from vector $\mathbf{u}$. Let us call such sets *equidistant classes*. It proves that only equidistant classes can serve as the ground state of the net: under particular conditions all configurations of one class (and no other) provide a global minimum to the energy functional. The composition and the number of equidistant classes are defined by vector $\mathbf{u}$. The conditions that make one or another class become the ground state are determined by $f(x)$ and $g(x)$.

The possibility to make the ground state multiply degenerate by choosing vector $\mathbf{u}$ is a valuable advantage of the approach. The ground state can hold a great deal of configurations: the number of configurations is a polynomial function of the dimensionality p. That is to say, it becomes possible to build networks of very large memory.

Litinskii L. and Malsagov M..
Multiplicative Neural Network with Thresholds.
DOI: 10.5220/0004629605230528
In *Proceedings of the 5th International Joint Conference on Computational Intelligence* (NCTA-2013), pages 523-528
ISBN: 978-989-8565-77-8

The disadvantage of the method is that not any set of configuration can serve as the ground state. This state can't consist of fully random configurations because the configurations must be equally distant from vector $\mathbf{u}$. Equidistant configurations are located around vector $\mathbf{u}$ symmetrically. And that is the limitation of the whole approach. How we can overcome this restriction is considered at the end of the paper.

In the next section we give the main results of the work and their short explanations, and consider one specific example. In the final section we analyze the potentials and limitations of the approach.

2 MAIN RESULTS

The energy of state $\mathbf{s}$ of network (1) is equal to

$$E(\mathbf{s}) \sim -\left(\sum_{i,j=1}^{p} J_{ij} s_i s_j + 2\sum_{i=1}^{p} T_i s_i \right) = \\ = -\left(f(x)(\mathbf{u},\mathbf{s})^2 + 2g(x)(\mathbf{u},\mathbf{s}) \right),$$

where $(\mathbf{u},\mathbf{s})$ is the scalar product of p-dimensional normalized vectors $\mathbf{u}$ and $\mathbf{s}$: $(\mathbf{u},\mathbf{s}) = \sum_{i=1}^{p} u_i s_i$. In further consideration it will be better to seek maxima of $F(\mathbf{s}) = -E(\mathbf{s})$:

$$F(\mathbf{s}) = f(x)(\mathbf{u},\mathbf{s})^2 + 2g(x)(\mathbf{u},\mathbf{s}) \to \max \qquad (2)$$

2.1 Classes Σ_k

Functional $F(\mathbf{s})$ takes the same value for all configurations the scalar products of which by vector $\mathbf{u}$ have the same value. Let us introduce the cosine of the angle between vectors $\mathbf{s}$ and $\mathbf{u}$:

$$\cos w = (\mathbf{s},\mathbf{u}) / p.$$

When $\mathbf{s}$ runs over 2^p possible configurations, $\cos w$ doesn't necessarily takes 2^p different values. Let us number different values of the cosine in descending order starting the numbering with 0:

$$\cos w_0 > \cos w_1 > ... > 0 > ... > \cos w_{t-1} > \cos w_t. \qquad (3)$$

The number $t+1$ of different values of the cosine does not exceed 2^p. Let Σ_k stand for the class of configurations $\mathbf{s}$ such that the cosine of the angle between $\mathbf{s}$ and vector $\mathbf{u}$ is $\cos w_k$:

$$\Sigma_k = \{ \mathbf{s} : (\mathbf{s},\mathbf{u}) = p \cdot \cos w_k \}, \ k = 0,1,...,t. \qquad (4)$$

Clear that each configuration from class Σ_k is the same distance away from vector $\mathbf{u}$, other configurations being a different distance off $\mathbf{u}$.

We see that functional $F(\mathbf{s})$ (2) takes $t+1$ values no matter what value x takes. All we have to do to find the ground state is to find the greatest among $t+1$ values:

$$F_k \sim f(x)\cos^2 w_k + 2\frac{g(x)}{p}\cos w_k, \ k = 0,1,..,t. \qquad (5)$$

The number of classes Σ_k and their composition are determined by vector $\mathbf{u}$ solely. With that, F_k are determined by cosines $\cos w_k \sim (\mathbf{s},\mathbf{u})$ for fixed f, g and x. We restrict our consideration to the case when vectors $\mathbf{u}$ have only nonnegative coordinates. The results can be easily extended to the case when some of u_i are negative (see below). We will assume that u_i are arranged in ascending order:

$$0 \le u_1 \le u_2 \le ... \le u_p.$$

It is easy to see that the sequence of cosines (3) is symmetric about its middle point:

$$\cos w_k = -\cos w_{t-k}, \ \Sigma_k = -\Sigma_{t-k}, \quad \forall k \le t.$$

If the number of different classes is even ($t+1 = 2l$), the cosines first go down to their positive minimum $\cos w_{l-1}$, then they become negative:

$$\cos w_0 > \cos w_1 > ... > \cos w_{l-1} > 0 > \\ > \cos w_l = -\cos w_{l-1} > ... > \cos w_t = -\cos w_0.$$

None of the cosines of the sequence is zero. On the other hand, when $t = 2l$, one of the cosines (3) is zero, and the sequence has the form:

$$\cos w_0 > ... > \cos w_{l-1} > \cos w_l = 0 > \\ > \cos w_{l+1} = -\cos w_{l-1} > ... > \cos w_t = -\cos w_0.$$

In this case Σ_l-class configurations are orthogonal to vector $\mathbf{u}$.

By way of example let us build a few starting classes Σ_k when the coordinates of vector $\mathbf{u}$ obey the following rule: $0 < u_1 < u_2 = u_3 < u_4 < ... < u_p$ with $2u_2 < u_4$. Class Σ_0 holds configurations that are nearest to vector $\mathbf{u}$, so $\Sigma_0 = \{\mathbf{e}\}$, where $\mathbf{e} = (1,1,..,1)$. The corresponding cosine is equal to $\cos w_0 = \sum_1^p u_i / p$. Class Σ_1 consists of configurations that are a bit more distant from $\mathbf{u}$ than Σ_0-class configurations. In our case it gives $\Sigma_1 = \{(-1,1,...,1)\}$, and $\cos w_1 = \cos w_0 - 2u_1/p$. The next class holds two configurations $(1,-1,1,..,1)$ and $(1,1,-1,..,1)$, and $\cos w_2 = \cos w_0 - 2u_2/p$. Class Σ_3 also consists of two configurations $(-1,-1,1,..,1)$ and $(-1,1,-1,..,1)$, and $\cos w_3 = \cos w_0 - 2(u_1+u_2)/p$. Class Σ_4 holds one configuration $\Sigma_4 = \{(1,-1,-1,1,...,1)\}$, and $\cos w_4 = \cos w_0 - 4u_2/p$. So does class Σ_5: $\Sigma_5 = \{(1,1,1,-1,...,1)\}$, $\cos w_5 = \cos w_0 - 2u_4/p$. And so on. To distribute all configurations into classes Σ_k, it is necessary to arrange in ascending order all 2^p possible sums $\sum_{i=1}^{p} \varepsilon_i u_i$ where coefficient ε_i can be either 0 or 1. This task is similar to the number partitioning problem (Mertens, 2001). In our case it is not necessary to try to solve the problem in general.

Another example. It is not difficult to describe the distribution of configurations among equidistant classes when vector $\mathbf{u} = \mathbf{e} = (1,1,...,1)$. It is easy to see that in this case the cosines take $p+1$ different values:

$$\cos w_k = 1 - 2k/p,\ k = 0,1,...,p, \tag{6}$$

and the k-th class holds the configurations that have exactly k negative coordinates. Let us introduce a special notification for such classes:

$$\Sigma_k^{(\mathbf{e})} = \left\{ \mathbf{s} : (\mathbf{s},\mathbf{e}) = \sum_{i=1}^{p} s_i = p - 2k \right\}, k = 0,1,...,p. \tag{7}$$

The number of configurations in class $\Sigma_k^{(\mathbf{e})}$ is $\binom{p}{k}$.

Further one or another $\boldsymbol{\sigma}$-configuration will be often used as vector $\mathbf{u}$. Basing on classes $\Sigma_k^{(\mathbf{e})}$ it is simple to understand the structure of equidistant classes in this case. Clear that both the number of different cosines and their values remains the same as with $\mathbf{u} = \mathbf{e}$ (see (6)). Coordinatewise multiplication of all configurations from class $\Sigma_k^{(\mathbf{e})}$ by $\boldsymbol{\sigma}$-configuration is used to obtain class $\Sigma_k^{(\boldsymbol{\sigma})}$ from class $\Sigma_k^{(\mathbf{e})}$ (7):

$$\Sigma_k^{(\boldsymbol{\sigma})} = \left\{ \mathbf{s} : (\mathbf{s},\boldsymbol{\sigma}) = \sum_{i=1}^{p} s_i \sigma_i = p - 2k \right\}, k = 0,1,...,p.$$

2.2 Functions f(x) and g(x)

Now let us consider the role of functions $f(x)$ and $g(x)$. Collection of $\{F_k(x)\}_0^t$ (5) is a family of functions of x. Function $F_l(x)$, which surpasses other functions at particular x, determines ground-state class Σ_l.

Let the amplitude of function $F_l(x)$ at point x_0 be greater than amplitudes of other functions: $F_l(x_0) > F_k(x_0)\ \forall k \neq l$. If $f(x)$ and $g(x)$ are continuous functions, a small variation of x does not change the superiority of $F_l(x)$ over other functions in the general case. Class Σ_l keeps being the ground state in a small vicinity of x_0. If x changes on, it becomes almost inevitable that function $F_l(x)$ intersects another function, say, $F_n(x)$. After that it is $F_n(x)$ that starts exceeding all other functions. At the point of intersection of functions the ground state passes to class Σ_n: the transition of the ground state $\Sigma_l \to \Sigma_n$ takes place. Of course, the transition point is defined by forms of functions $f(x)$, $g(x)$ and cosines (3). However, something about the way the ground state changes can be understood from the general considerations.

Let us rearrange formula (5) by taking $f(x)$ out of the brackets and completing the expression in the brackets to the square. Accurate to insignificant items, the formula we get is

$$F_k \sim f(x)\left[\cos w_k - \gamma(x)\right]^2, \quad \gamma(x) = -\frac{g(x)}{p \cdot f(x)}, k = 0,1,..,t. \tag{8}$$

Let us first assume that $f(x) > 0$. In this event it is necessary to maximize the modulus of the bracketed expression with respect to k to find the largest F_k:

$$\max_k \left|\cos w_k - \gamma(x)\right|. \tag{9}$$

If $\gamma(x)$ is negative, the maximum of modulus (9) is ensured by the greatest value of the cosine, and the solution of (9) is $k = 0$. In this case, the ground state is associated to class Σ_0. Conversely, if $\gamma(x)$ is positive, the maximum of modulus (9) is ensured by the smallest value of the cosine. The solution of (9) is $k = t$ in this event, and the ground state is attributed to class $\Sigma_t = -\Sigma_0$. So, when $f(x)$ is positive, either class Σ_0 (if $g(x) > 0$) or class Σ_t (if

$g(x) < 0$) becomes the ground state.

Let us now examine what happens if $f(x) < 0$. In this case it is necessary to minimize the modulus of the bracketed expression (8) with respect to k to find the largest F_k. Generally speaking, to do it is not at all difficult: it is just necessary to define $\cos w_k$ that is closest to the current value of $\gamma(x)$. The corresponding class Σ_k will be the ground state of the net. Let us look at Figure 1 to understand collisions that occur in this case.

In Figure 1 the Y-axis carries representative values of $\cos w_k$ for $k = l-1, l$ and $l+1$. The steadily decreasing curve represents function $\gamma(x)$. c_k denotes the half sum of two successive values of the cosine:

$$c_k = \frac{\cos w_{k-1} + \cos w_k}{2}, \quad k = 1, 2, ..., t. \qquad (10)$$

The value of x at which $\gamma(x) = c_k$ is indicated as x_k:

$$\gamma(x_k) = c_k \Leftrightarrow x_k = \gamma^{-1}(c_k). \qquad (11)$$

Let x belong to interval (x_l, x_{l+1}) initially: $x_l < x < x_{l+1}$. It is easy to see that for any x from this interval it is $\cos w_l$ that is nearest to $\gamma(x)$. So, $k = l$ is the solution of (9), and class Σ_l serves as the ground state of the net. Note that it is true for all x in the interval (x_l, x_{l+1}). Variable x can grow (fall) until it steps over x_{l+1} (x_{l-1}) and the ground state passes to class Σ_{l+1} (Σ_{l-1}), and so on.

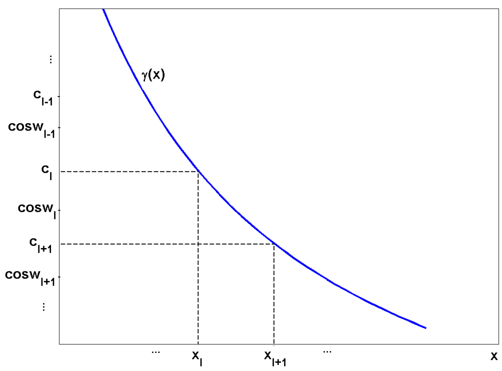

Figure 1: Graphical solution of the problem (9): see body of the paper.

We see that when $f(x) < 0$ and $\gamma(x)$ is a continuous function, the changing of the ground state changes its number by 1: $\Sigma_k \to \Sigma_{k\pm1}$. There is a kind of continuity in its number changing with parameter x. In principle, it is possible to organize "discontinuous" control over ground-state "jumps" $\Sigma_k \to \Sigma_l$ so that class numbers k and l would differ by more than 1. For this purpose one should use either discontinuous function $g(x)$, or the fact that when $f(x)$ becomes positive, the ground state passes from any class Σ_k to either class Σ_0 or Σ_t.

2.3 Example

To exemplify the results let us consider functions $f(x)$ and $g(x)$ of the following form (Litinskii, 1999):

$$f(x) = 1 - 2x, \; g(x) = q(1-x), \; q \geq 1.$$

In this case $F_k(x)$ in (5) takes the form:

$$F_k(x) = \left(q + p\cos w_k\right)^2 - 2xqp\cos w_k\left(q + p\cos w_k\right).$$

Competing functions $F_k(x)$ are a family of straight lines whose structure can be examined easily. As a result, we get the following statement.

Theorem. *When x grows indefinitely from the initial value of 0, the ground state of a net passes consecutively to classes Σ_k (4): $\Sigma_0 \to \Sigma_1 \to \Sigma_2 \to ... \to \Sigma_{k_{\max}}$. Transition $\Sigma_{k-1} \to \Sigma_k$ occurs at critical point*

$$x_k = \frac{q/p + \left(\cos w_{k-1} + \cos w_k\right)/2}{q/p + \cos w_{k-1} + \cos w_k}, \; k = 1, 2, ..., k_{\max},$$

and as long as $x \in (x_k, x_{k+1})$, class Σ_k is the ground state of the net. Number $k_{\max}$ of the last transition is determined by the requirement that denominator $q/p + \cos w_{k-1} + \cos w_k$ should be positive. If vector $\mathbf{u}$ *is configuration, the ground-state configurations are the only fixed points of the net.*

The composition of classes Σ_k is not detailed in the theorem at all: classes consist of configurations equally distant from vector $\mathbf{u}$. After classes Σ_k are defined with the aid of $\mathbf{u}$, change of parameter x results in the ground state jumping from one class to another. It is possible to show that independently of vector $\mathbf{u}$ the first transition of the ground state $\Sigma_0 \to \Sigma_1$ occurs after ½: $x_1 > 1/2$. Additionally, it turns out that $k_{\max}$ is always greater than $p/2$, and $x_{k_{\max}} > 1$. The use of factor q makes it possible to regulate the total number of ground-state transitions.

3 DISCUSSION AND CONCLUSIONS

The findings from the previous paragraph allow us to control the ground state of the net to a considerable extent. Let us consider a p-dimensional hyper-cube with edge length of 2 and center at the origin of coordinates. Configurations $\mathbf{s}$ are located at cube vertexes. Symmetric directions in the hyper-cube must be chosen as vector $\mathbf{u}$. For each $\mathbf{u}$ of that kind 2^p $\mathbf{s}$-configurations are distributed in symmetric sets with vector $\mathbf{u}$ being the axis of symmetry. Each set like that forms one of Σ_k classes. It can be turned into the ground state by using the approach offered. Particularly, it is possible to create the ground state from a very large number of configurations. For example, the number of $\Sigma_k^{(\mathbf{e})}$-class configurations (7) is equal to $p!/\left(k!(p-k)!\right)$.

Some coordinates of vector $\mathbf{u}$ can be zero. Let $u_1 = 0$. Then the same class will comprise not only configuration $\mathbf{s} = (s_1, s_2, ..., s_p)$, but also configuration $\mathbf{s}' = (-s_1, s_2, ..., s_p)$. In other words, vector $\mathbf{u}$ having a zero coordinate results in the number of configurations doubling in each class Σ_k. In this event the conclusive statement of Theorem is more general and should read: if non-zero coordinates of vector $\mathbf{u}$ are equal to each other, the ground-state configurations are the only fixed points of the net.

What possible consequences the approach can have are not known yet. It is necessary to look through all symmetric directions of $\mathbf{u}$ in the hyper-cube and arrange cube vertexes with respect to vertex-to-vector $\mathbf{u}$ distance in each case. It is necessary to turn to methods of the group theory here (Davis, 2007).

The disadvantage of the whole approach is that configurations comprising the ground state can't be arbitrary. They are the same distance from vector $\mathbf{u}$ and, therefore, form a symmetric set. We hope that the following tricks (or their combinations) can help us to avoid total symmetry of the ground state. First, we can use a few vectors like $\mathbf{u}$ into the connection matrix and thresholds rather just one vector. For example, let there be vector $\mathbf{v} = (v_1, v_2, ..., v_p)$, $\|\mathbf{v}\|^2 = p$, and let us consider a neural net similar to (1):

$$\begin{cases} J_{ij} = f(x)\left((1-\delta_{ij})(u_i u_j + v_i v_j)\right), \\ T_i = g(x)(u_i + v_i), \\ s_i(\tau+1) = \operatorname{sgn}\left(\sum_{j=1}^{p} J_{ij} s_j(\tau) + T_i\right). \end{cases} \tag{12}$$

If vectors $\mathbf{u}$ and $\mathbf{v}$ are configurations, it proves that as long as x does not exceed the first transition point x_1, the initial configurations $\mathbf{u}$ and $\mathbf{v}$ themselves are the ground state. If $x > x_1$, a set of configurations equally distant from both $\mathbf{u}$ and $\mathbf{v}$ will constitute the ground state. The net (12) will not have other fixed points. Supported by a computer simulation, this result arouses cautious optimism.

Second, it is possible to "separate" in (1) thresholds T_i and numbers u_i used for building the multiplicative matrix M_{ij}. Let us use earlier-introduced vector $\mathbf{v}$ and consider a neural net

$$\begin{cases} J_{ij} = f(x)(1-\delta_{ij})u_i u_j, \quad T_i = g(x) v_i, \\ s_i(\tau+1) = \operatorname{sgn}\left(\sum_{j=1}^{p} J_{ij} s_j(\tau) + T_i\right). \end{cases}$$

Tentative considerations show that its ground state is formed by Σ_k-class configurations nearest to the vector difference $\mathbf{u} - \mathbf{v}$. In other words, the trick allows us to avoid the total symmetry of the ground state. Of course the results need closer research.

The memory of the standard Hopfield model with the Hebbian connection matrix and random and independent patterns $\mathbf{s}^{(\mu)}$ is well understood. However, if the connection matrix is of the general form, the memory of such a network is practically unknown. In the same time an arbitrary connection matrix $\mathbf{J}$ can be presented as a quasi-Hebbian one, when using: *i*) orthogonal vectors $\mathbf{u}^{(\mu)}$ related to the eigenvectors of the matrix $\mathbf{J}$,

$$J_{ij} = (1-\delta_{ij})\sum_{\mu=1} u_i^{(\mu)} u_j^{(\mu)} \;\Leftrightarrow\; \mathbf{J} \sim \sum_{\mu=1} \mathbf{u}^{(\mu)+}\mathbf{u}^{(\mu)}$$

where $\mathbf{u}^{(\mu)} = (u_1^{(\mu)}, ..., u_p^{(\mu)})$, $u_i^{(\mu)} \in \mathbf{R}^1$, $\mathbf{u}^{(\mu)} \perp \mathbf{u}^{(\nu)}$, *ii*) or configuration vectors $\mathbf{s}^{(\mu)}$ with the weights r_μ (Kryzhanovsky, 2007):

$$\mathbf{J} \sim \sum_{\mu=1} r_\mu \mathbf{s}^{(\mu)+}\mathbf{s}^{(\mu)}, \; s_i^{(\mu)} = \pm 1, \; r_\mu \in \mathbf{R}^1.$$

Our multiplicative matrix $\mathbf{M}$ is only one term of the quasi-Hebbian expansion. We hope that a detailed analysis of the network with the connection matrix $\mathbf{M}$ will allow us to make headway on investigating a more general case.

ACKNOWLEDGEMENTS

The work was supported by Russian Basic Research Foundation (grants 12-07-00259 and 13-01-00504).

REFERENCES

Hertz, J., Krogh, A., Palmer, R., 1991. *Introduction to the Theory of Neural Computation.* Massachusetts: Addison-Wesley.

Mertens, S., 2001. *A Physicist's Approach to Number Partitioning.* Theoret. Comput. Sci. 265, 79-108.

Litinskii, L. B., 1999. *High-symmetry Hopfield-type neural networks.* Theoretical and Mathematical Physics, 118: 107-127.

Davis, M. W., 2007. *The geometry and topology of Coxeter groups.* Princeton: Princeton University Press.

Kryzhanovsky, B. V., 2007. *Expansion of a matrix in terms of external products of configuration vectors.* Optical Memory & Neural Networks (Information Optics), 16: 187-199.