Connectivity Maximization of Layered Neural Networks for Supervised Learning

E. Fiesler

IDIAP, Case postale 609, CH-1920 Martigny, Switzerland

Abstract

One of the main problems in current artificial neural network engineering is the lack of design rules for layered neural network topologies, namely how many hidden layers and how many neurons per hidden layer to choose for a neural network. This paper offers a theoretical basis for approaching this problem. Formally proven theories are developed which maximize the interconnection topology of layered neural networks, which use supervised learning, to obtain a maximum number of interconnections and therefore allow a maximum potential storage capacity. The results presented here depend only on the neural network statics and are therefore learning rule independent.

Keywords: (artificial) neural network, connectionism, neural network topology, neural network statics, neural network connectivity, neural network architecture, neural network capacity, neural network taxonomy, high(er) order neural network, Sigma-Pi neural network

1 Introduction

Although the field of artificial neural networks is a rapidly growing one, some fundamental questions remain unanswered, like how to configure a layered neural network. More specifically, there is a lack of configuration rules for choosing the number of hidden layers and the number of neurons for each of those hidden layers.

Since neural networks process and store information, their (storage) capacity is a key design issue, and since the interconnection strengths (or *weights*) contain the information of the neural network, the (storage) capacity is proportional to the total number of weights in the network. Without a-priori knowledge, all the weights of a neural network are considered to have the same potential, that is, the range of their possible values is identical. A fully connected neural network will therefore have a higher potential capacity than any other interconnection scheme. However, the maximum number of connections depends on the topology of the neural network.

Most neural networks are trained using a supervised learning rule. For these networks, the number of input and output neurons is determined by the application and are therefore considered constants in this publication. The total number of neurons in the network (N) is, without loss of generality, also considered a constant, in order to enable comparisons between different interconnection schemes.

For a description of neural network statics and related terminology see [Fiesler-92], and for the proofs of all the theorems plus a more extensive bibliography, see [Fiesler-94].

2 Maximal Interconnection topologies

Depending on the types of interconnections present, the number of connections in a supervised neural network can be maximized by varying the topology of the network.

[0]A compact version of this paper has been published in ICANN '93; Proceedings of the International Conference on Artificial Neural Networks, S. Gielen and B. Kappen (editors), p. 514. Springer-Verlag, London, UK, 1993. ISBN: 0-387-19839-3.

2.1 Interlayer connections

Fully 'interlayer connected' neural network topologies are the most interesting because they are used in many neural network paradigms. For these networks, the total number of weights (W) is $\sum_{l=2}^{L} N_{l-1}\, N_l$ (from [Fiesler-90]), where N_l is the number of neurons in layer l. The smallest neural network with interlayer connections is a two layer network. This network has no free parameters since the size of both the input and the output layer are fixed. A three layer network leaves one variable: the size of the hidden layer, which is given by $N - N_1 - N_3$. The first interesting architecture is therefore a four layer one. Here we have: $L = 4$ and $W(N_2, N_3) = N_1 N_2 + N_2 N_3 + N_3 N_4$. Since $N_2 = N - N_1 - N_3 - N_4$, N_3 can be eliminated from $W(N_2, N_3)$, leaving $W(N_2) = (N - N_1 - N_2 - N_4)(N_2 + N_4) + N_1 N_2$. To maximize $W(N_2)$, the partial derivative of $W(N_2)$ with respect to N_2 has to be determined:

$$\frac{\partial W(N_2)}{\partial N_2} = N - 2N_2 - 2N_4.$$

By setting this quantity equal to zero, a (global) maximum of $\frac{N^2}{4} - N_1 N_4$ is found at $N_2 = \frac{N}{2} - N_4$. Analogously, by eliminating N_2, one obtains $N_3 = \frac{N}{2} - N_1$. Note the interesting interdependence between layers 1 and 3, and 2 and 4.

Under the constraints of keeping N_1 and N_L constant, the maximal interconnection topologies for any number of layers can be determined using the same technique. Table 1 gives

L	N_2	N_3	N_4
2	N_L	-	-
3	$N - N_1 - N_L$	N_L	-
$4k$	$\frac{2N}{L} - N_L$	$\frac{2N}{L} - N_1$	N_L
$4k+1$	$\frac{2(N-N_1)}{L-1} - N_4$	$\frac{2(N-N_1)}{L-1} - N_1$	$\frac{2(N-N_1)}{L-1} - N_2$
$4k+2$	N_L	$\frac{2(N-N_L)-LN_1}{L-2}$	$\frac{2(N-N_1)-LN_L}{L-2}$
$4k+3$	$N - \frac{L-1}{2}(N_1 + N_L)$	N_L	$\frac{L+1}{2}(N_1 + N_L) - N$

L	$W_{\max}$	*constraints*
2	$N_1 N_L$	$N_1, N_L < N$ and $N_1 + N_L = N$
3	$(N - N_1 - N_L)(N_1 + N_L)$	$N_1 + N_L < N$
$4k$	$\frac{N^2}{L} - N_1 N_L$	$N_1, N_L < \frac{2N}{L}$
$4k+1$	$\frac{(N-N_1)^2}{L-1}$	$N_1 = N_L < \frac{2N}{L+1}$
$4k+2$	$\frac{(N-N_1-N_L)^2}{L-2} + N_1 N_L$	$\frac{L}{2}N_1 + N_L < N$ and $N_1 + \frac{L}{2}N_L < N$
$4k+3$	$(N_1 + N_L)(N - \frac{L+1}{4}(N_1 + N_L))$	$\frac{2N}{L+1} < N_1 + N_L < \frac{2N}{L-1}$

Table 1: Maximal interconnection topologies and maximum number of weights for fully interlayer connected neural networks.

an overview of the maximal interconnection topologies for any size neural network with interlayer connections only. For networks with more than three layers, the number of layers is expressed as a four-fold ($4k$, where $k \in \mathcal{Z}^+$) plus a constant. The last column of the second part of table 1 shows the constraints on the size of the input and output layer. Note that the two layer solution is only valid under the condition that $N = N_1 + N_2$. If this is not the case, the maximum number of weights is obtained by either a three or a four layer network. The precise conditions are given by this theorem:

Theorem 1 *The maximum interconnection topology for a fully interlayer connected neural network with a fixed input and output layer size is : if $N_1 + N_L = N$: a two layer neural network; or, if $N_1^2 + N_L^2 + N_1 N_L + \frac{N^2}{4} > N(N_1 + N_L)$: a three layer neural network with $N_2 = N - N_1 - N_3$; otherwise a four layer network with $N_2 = \frac{N}{2} - N_4$, and $N_3 = \frac{N}{2} - N_1$. The total number of weights for each of the maximum interconnection topologies is presented in the second part of table 1.*

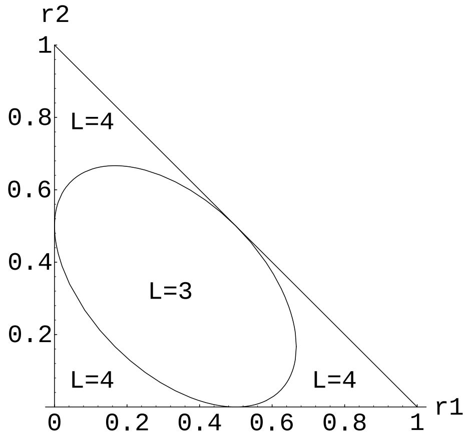

Figure 1: Number of layers of the maximal interconnection topologies for fully interlayer connected neural networks.

(The decision criteria given in theorem 1 are obtained by comparing their respective '$W_{\max}$'s'.) Figure 1 shows graphically how to choose the maximal interconnection topology (`r1` $= N_1/N$ and `r2` $= N_L/N$).

2.2 Interlayer and intralayer connections

For neural networks with both interlayer and intralayer connections, the maximal interconnection topologies can be determined analogously. For these networks, only one or two layer systems are optimal. For more than two layers the outcome of the optimization is: drop all but one or two layers and distribute the neurons over these layers. This can be observed by comparing the formulas in the $W_{\max}$ column in the second part of table 2 which shows that increasing the number of layers decreases the maximum number of weights. One and two layer neural networks which have all possible interlayer and intralayer connections (including self-connections), are equivalent to a plenary neural network[1] with the same number of neurons. The maximum number of weights[2] $W_{\max} = \frac{N}{2}(N \pm 1)$ is therefore the same as for a plenary neural network.

Theorem 2 *The maximum interconnection topology for a fully interlayer and intralayer connected neural network with given input and output layer size is one with one or two layers, and $W_{\max} = \frac{N(N \pm 1)}{2}$.*

[1] A *plenary neural network* is a neural network with all possible inter-, intra-, and supralayer connections.

[2] The $\pm$-symbol denotes the option of having self-connections. If self-connections are present, addition has to be used, and subtraction otherwise.

	maximal interconnection topology	
L	N_2	N_3
1	-	-
2	N_L	-
$3k$	$\frac{3N}{L} - N_1 - N_L$	N_L
$3k+1$	$\frac{3N-(L+2)N_1}{L-1} - N_3$	$\frac{3N-(L+2)N_1}{L-1} - N_2$
$3k+2$	N_L	$\frac{3N-(L+1)(N_1+N_L)}{L-2}$

L	$W_{\max}$	*constraints*
1	$\frac{N^2}{2} \pm \frac{N}{2}$	$N_1 = N_L$
2	$\frac{N^2}{2} \pm \frac{N}{2}$	$N_1, N_L < N$ and $N_1 + N_L = N$
$3k$	$\frac{3N^2}{2L} - N_1 N_L \pm \frac{N}{2}$	$N_1 + N_L < \frac{3N}{L}$
$3k+1$	$\frac{3N^2-6N_1N+(L+2)N_1^2}{2(L-1)} \pm \frac{N}{2}$	$N_1 = N_L < \frac{3N}{L+2}$
$3k+2$	$\frac{3N^2-6N(N_1+N_L)+(L+1)(N_1+N_L)^2}{2(L-2)} \pm \frac{N}{2}$	$N_1 + N_L < \frac{3N}{L+1}$

Table 2: maximal interconnection topologies and maximum number of weights for fully interlayer and intralayer connected neural networks.

2.3 Interlayer and supralayer connections

Since two layer neural networks do not have supralayer connections, the smallest networks with both interlayer and supralayer connections are three layer networks. As observed before, three layer networks have no degrees of freedom in choosing their topology, since $N_2 = N - N_1 - N_3$. The number of weights is $(N - N_1 - N_3)(N_1 + N_3) + N_1 N_3$ ([Fiesler-90]). Optimization, using a similar technique as outlined in section 2.1, shows that the larger the number of layers, the larger is the maximum number of weights. This leads to the following theorem:

Theorem 3 *The maximal interconnection topology for fully interlayer and supralayer connected neural networks with a fixed input and output layer size is three or more layers with the neurons equally divided over the layers (that is, $N_l = \frac{N-N_1-N_L}{L-2}$ for all $1 < l < L$). This gives a total number of weights of $\frac{(L-3)N^2+2N(N_1+N_L)+(1-L)(N_1^2+N_L^2)-2N_1N_L}{2(L-2)}$.*

Note that

$$\lim_{L\to\infty} \frac{(L-3)N^2 + 2N(N_1+N_L) + (1-L)(N_1^2+N_L^2) - 2N_1N_L}{2(L-2)} = \frac{N^2 - N_1^2 - N_L^2}{2}, \qquad (1)$$

which gives the theoretical maximum of $\frac{N^2}{2}$ for $N_1 = N_L = 0$.

2.3.1 Plenary neural networks

Since all plenary neural networks are topologically identical, the following theorem holds:

Theorem 4 *The topology of all plenary neural networks is maximally interconnected, and the maximum number of weights is $\frac{N(N+1)}{2}$.*

Proof The proof is trivial, since the topology of a plenary neural network is equivalent to a fully connected (undirected) graph, whose number of edges only depends on the number of vertices. □

3 Summary

For each of the interconnection schemes the maximal interconnection topology is obtained and presented in the form of a series of analytically proved theorems[3]. The maximal interconnection topology and maximum number of weights for all the interconnection schemes for fully connected layered neural networks with given (fixed) input and output layer sizes are given in table 3, where $L_{\min}$ represents the minimum number of layers, and $L_{\max}$ the maximum number of layers.

interconnection scheme	*maximal interconnection topology*			$W_{\max}$
	$L_{\min}$	$L_{\max}$	N_l's	
inter ($L = 3$)	3	3	$N_2 = N - N_1 - N_L$	$(N - N_1 - N_L)(N_1 + N_L)$
inter ($L = 4$)	4	4	$N_2 = \frac{2N}{L} - N_L$, $N_3 = \frac{2N}{L} - N_1$	$\frac{N^2}{L} - N_1 N_L$
inter & intra	1	2	fixed distribution	$\frac{N(N\pm1)}{2}$
inter & supra	3	no max.	$\forall l : N_l = \frac{N-N_1-N_L}{L-2}$	see theorem 3
plenary	1	no max.	any distribution	$\frac{N(N\pm1)}{2}$

Table 3: maximal interconnection topologies and maximum number of weights for all fully connected interconnection schemes.

References

[Fiesler-92] E. Fiesler, **Neural Network Formalization**, IDIAP Technical report 92-01, IDIAP, Martigny, Switzerland, July 1992; also submitted to *Computer Standards & Interfaces*, special issue on Neural Network Standards, John Fulcher, editor, North-Holland, Elsevier Science Publishers B. V., Amsterdam, The Netherlands..

[Fiesler-94] E. Fiesler, H. J. Caulfield, A. Choudry, and J. P. Ryan, **Maximal Interconnection Topologies for Neural Networks**, to be submitted to *IEEE Transactions on Neural Networks*.

[Fiesler-90] E. Fiesler, H. J. Caulfield, A. Choudry, and J. P. Ryan, **Maximum Capacity Topologies for Fully Connected Layered Neural Networks with Bidirectional Connections**, *Proceedings of the International Joint Conference on Neural Networks (IJCNN) San Diego*, volume I, pages 827–831, San Diego, California, June 17-21, IEEE Neural Networks Council; Edward Brothers, Ann Arbor, Michigan, 1990, IEEE Catalog number: 90CH2879-5.

[3]Note that although the presented results guarantee maximum interconnection topologies, this does not necessarily imply that the superposition of a specific learning rule will not influence these results.