A Novel Spatial Architecture Artificial Neural Network Based on Multilayer Feedforward Network with Mutual Inhibition among Hidden Units

Gang Yang\(^1\), Junfei Qiao\(^1\), and Mingzhe Yuan\(^2\)

\(^1\) Intelligent Systems Institute, College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing, 100124, China
\(^2\) Shenyang Institute of Automation Chinese Academy of Science, Shenyang, 110016, China

\textit{hank.yang2010@gmail.com, isibox@sina.com}

\textbf{Abstract.} We propose a Spatial Artificial Neural Network (SANN) with spatial architecture which consists of a multilayer feedforward neural network with hidden units adopt recurrent lateral inhibition connection, all input and hidden neurons have synapses connections with the output neurons. In addition, a supervised learning algorithm based on error back propagation is developed. The proposed network has shown a superior generalization capability in simulations with pattern recognition and non-linear function approximation problems. And, the experimental also shown that SANN has the capability of avoiding local minima problem.

\textbf{Keywords:} multilayer feedforward network, spatial span connection, lateral inhibition mechanism.

\section{Introduction}

Warren McCulloch and Walter Pitts [1] gave an academic background of applying artificial neural networks (ANNs) because they have shown that ANNs have the capability of computing any arithmetic or logical function in principle. Because of the characteristic of resembling the process we go through learning something [2], good at providing fast and close approximations of the correct answer, and best at identifying patterns and trends in data, ANNs have already been successfully applied in many industries and daily life for prediction and forecasting [3-8].

Theoretical research of ANNs major focus on two points: architecture (BP, Hopfield network, RBF etc.) and learning algorithm (Back propagation, Levenberg-Marquardt, Conjugate Gradient algorithm etc.). These structures of ANNs basically are planar and symmetrical connection. But the animal’s neuronal system is there dimensional and asymmetry; and the neurons of brain in previous layers have narrow perception field but have wider perception fields in deep layers. On the other hand, structure defines function in ANNs. So, there
are some structure studies, for instance, based on functional expansion, Chao proposed a new pyramid network [9] and analyzed the representation and generalization theoretically and experimentally; Jaeger presented Echo State Networks [10], which combined the trainable linear combination of nonlinear response signals of internal units and the desired output signals. Research of cortex shown that the inhibition of cerebral cortex may avoid the excitatory activity caused by a stimulus ripple across the entire network and prevent the confused jumble resulted by the overlapping signals [11].

We propose a neural network with new spatial architecture called Spatial Artificial Neural Network (SANN) to tradeoff the structure and functions and aim to achieve higher representation and generalization capability. SANN is based on a multilayer feed-forward network, spatial span connection and recurrent lateral inhibitory connection mechanism. Through SANN, all non-output neurons have connections to the output neurons. Supervised training rules of SANN are then shown for weights update and learning.

The proposed network applied to several benchmark problems indicated that this network yields significantly representation capability and generalization performance, and it also can avoid the local minima problems in the supervised learning.

2 Definition of Spatial Artificial Neural Network

We define the spatial artificial neural network as follows, and the topology structure is shown in Fig. 1. The first part of SANN architecture is so-called the basic network lies in the center of SANN architecture. Actually the basic network is a multilayer feedforward network which adds mutual inhibition mechanism among the hidden units within a same layer. The feedforward connections among neurons are shown in Fig. 1 using arrowhead line, and its mutual inhibition adopts the recurrent lateral inhibition mechanism, as shown in Fig. 2. The second part of SANN is the spatial span connections part which is shown with arrowhead broken lines.

![Fig. 1. Architecture of SANN](image-url)
This basic network is used for transferring information from the environmental inputs to the network outputs. The recurrent lateral inhibitory connection of one hidden layer brings, which is widely found in biological neural systems, connection and competition to the hidden units.

The input signals were transferred to hidden units through the feedforward connections, and the outputs of hidden units corresponding to these input signals are called local output. And in the lateral inhibition theory, it is also called activation level. It will inhibit other hidden units with strength of inhibitory coefficients. The feedforward input combines with the inhibitory input (caused by other hidden units and calculated according to the product of the inhibiting neuron’s activation level and the corresponding recurrent lateral inhibitory coefficient) as the total input of hidden units, and the output was transferred to the next layers as the input signal.

The accessorional spatial span connections lay over the basic network of SANN. This part make that any two units of SANN in different layers may have direction connect according to the spatial span connection and weights. The global output of any unit through this span connection transfer signals to the next two or more layer neurons.

Both the inhibitory connection of hidden units and the spatial span connection between different layers guarantee the structure of SANN is a three-dimensional and asymmetry system, which resembles the biological neuronal system structure.

3 Mathematic Model of SANN

3.1 Review of Lateral Inhibition

Lateral inhibition (LI) is one type of inhibitions of cerebral cortex. It is one of the basic principles of information treatment within neural system and describes the capacity of an excited neuron to reduce the activity of its neighbor neurons.

In the design of our SANN network, the architecture in hidden layers uses the recurrent lateral inhibition to achieve competition.

The topology of lateral inhibitory connection without self-inhibition adapted in SANN is show in Fig. 2, and the mathematic model is given by [12]

\[ y_{o,j} = y_{i,j} - \sum_{r=1, r \neq j}^{h} v_{rj} (y_{o,r} - \theta_{rj}), \quad r = 1, 2, \ldots, h \]  

(1)

where \( h \) is the number of hidden units, \( y_{o,j} \) is the output of \( j \)-th hidden neuron, \( y_{i,j} \) is the environment stimulates of neuron \( j \) received, \( v_{rj} = v_{jr} \) is the lateral inhibitory coefficient between neuron \( r \) and \( j \), \( \theta_{rj} \) is the inhibiting threshold value of neuron \( j \) due to neuron \( r \).

3.2 Mathematic Model of SANN

Suppose the SANN has \( L+1 \) layers including the input and output layers, the \( l \)-th \((l = 1, 2, \ldots, L-1)\) hidden layer has \( n_l \) units. The input layer is denoted as
$l = 0$, has $n$ units; and the output layer is denoted as $l = L$, has $m$ units. Symbols $i$, $l$ and $k$ denote the input, hidden and output units respectively.

The output of the $k$-th unit in the output layer is given by:

$$o_k = f^L \left[ \sum_{l=1}^{L-1} \sum_{j=1}^{n_l} \omega_{jk} o_j^l + b_l \right] + \sum_{i=1}^{n} \omega_{ik} x_i + b_0], \quad k = 1, 2, \ldots, m \quad (2)$$

where the $f^l$ is the activation function, $o_j^l$ is the output of the $j$-th hidden neuron of the $l$-th layer, it is given by:

$$o_j^l = \begin{cases} f^l \left[ \sum_{i=1}^{n} \omega_{ij} x_i + b_0 \right] - \sum_{r=1}^{n_l} v_{rj} (o_r^l - \theta), & l = 1 \\ f^l \left[ \sum_{q=1}^{n_q} \omega_{pj} o_q^p + b_q \right] + \sum_{i=1}^{n} \omega_{ij} x_i + b_0] - \sum_{r=1}^{n_l} v_{rj} (o_r^l - \theta), & l \in [2, L - 1] \end{cases} \quad (3)$$

where $v_{rj}$ is the inhibitory coefficient from neuron $r$ to neuron $j$, $o_r^l$ is the local output of hidden unit $r$ in $l$-th layer, namely the activation level of neuron $r$, it is given by:

$$o_r^l = f^l \left[ \sum_{q=1}^{n_q} \omega_{pq} o_q^p + b_q \right] + \sum_{i=1}^{n} \omega_{ij} x_i + b_0], \quad p = 1, 2, \ldots, n_q \quad (4)$$

From Fig. 1 and Eq. (4) we can find that it is output of the current hidden neuron due to the feedforward input signals, namely the output comes from the input neurons. $v_{rj}$ is the recurrent lateral inhibitory weight between neuron $r$ and $j$ of hidden layer $l$.

4 Learning Algorithm

Here, we reference the standard back propagation algorithm to derivate the learning algorithm fit for SANN. It is described as follows.

Assume the three-layer SANN with $n$ input neurons, $h$ hidden neurons and one output neurons respectively. The output of the output unit $k$ due to the $p$-th ($p \in [1, P]$) input sample is given by $o_{pk}$, whereas the desired output is $t_{pk}$,
and the output of the \( j \)-th hidden unit for the \( p \)-th input pattern is given by \( o'_j^p \). Let \( \omega_{jk} \) be the weight between the \( k \)-th output unit and the \( j \)-th hidden unit, and \( \omega_{ij} \) be the weight between the \( j \)-th hidden unit and \( i \)-th input unit. The input for the \( i \)-th input unit due to the \( p \)-th input pattern is denoted by \( x_{pi} \).

According to the above definitions, the output of the \( j \)-th neuron in the hidden layer is given by:

\[
o'_j^p = f\left(\sum_{i=1}^{n} \omega_{ij} x_{pi} + \sum_{r=1}^{h} v_{rj} \bar{o}_{pr}\right)
\]  

(5)

where, \( \bar{o}_{pr} = f(\sum_{i=1}^{n} \omega_{ij} x_{pi}) \) is the output of the \( r \)-th hidden neuron due to the feedforward input signals, namely the output comes from the input neurons; \( v_{rj} \) is the recurrent lateral inhibitory weight between neuron \( r \) and \( j \) of hidden layer; \( f \) is the sigmoid activation function defined as: \( f(x) = 1/(1 + e^{-x}) \).

Similarly, the output of the \( k \)-th unit in the output layer is given by:

\[
o_k^p = f\left(\sum_{i=1}^{n} \omega_{ik} x_{pi} + \sum_{j=1}^{h} \omega_{jk} o'_j^p\right)
\]  

(6)

We define the sum of squared error of output unit with \( P \) samples as the system learning target function, to be:

\[
E = \frac{1}{2} \sum_{p=1}^{P} (t_{pk} - o_{pk})^2
\]  

(7)

The learning algorithm based error back propagation is to change the weights iteratively such that the function \( E \) Eq. (7) is minimized. The weight updates are proportional to \( \partial E \). According to the chain rule, we can obtain the partial derivative of \( \partial E \) with respect to \( w_{jk} \) and \( w_{ik} \), and the weight change for the \((t+1)\)-th iteration can be expressed as follows,

\[
\begin{align*}
\Delta \omega_{jk}(t+1) &= \mu \sum_{p=1}^{P} \delta_{pk} o'_j^p, \\
\Delta \omega_{ik}(t+1) &= \mu \sum_{p=1}^{P} \delta_{pk} x_{pi}, \\
\Delta \omega_{ij}(t+1) &= \mu \sum_{p=1}^{P} \delta_{pj} x_{pi},
\end{align*}
\]  

(8)

where the term \( \mu \) is the learning rate of the gradient method.

It is worthwhile to notice that the above gradient consists two parts, the first term is the basic back propagation in the basic network and the second term is the span back propagation in the spatial span connection network.

The update of SANN weights according to \( \omega_{sk}(t+1) = \omega_{sk}(t) + \alpha \Delta \omega_{sk} \). Where the term \( \alpha \) is the momentum term used to balance the stability and oscillate which may caused by the value selection of learning rate \( \mu \).

5 Numerical Experimental

We use two benchmark problems to compare the performance of SANN using the above supervised learning rule to train and the multilayer feed-forward network is trained by implementation of the back propagation algorithm in Matlab.
In our experiments, 10-fold cross validation is employed on each problem to compare the performance. The result is the average result of the ten folds.

5.1 Classification Capability (XOR Problem)

In order to consider the classification capability of SANN, we chose a simplest problem which is not linearly separable - “exclusive-or” problem which is also discussed in [13-15]. According to the complexity of XOR problem, in this section we consider to train 2-1-1 and 2-2-1 two architectures of SANN on the two dimensional XOR classification problem described in the first two rows of Table 1. Note that the inputs can still be 0 and 1 but the desired values must be changed keeping in mind the signal range.

(1) 2-1-1 SANN

Because only one hidden unit existed in the architecture of 2-1-1 SANN, there is no inhibitory effect or connection between the hidden units.

Set the value of global parameters are as follows: max iteration number $T = 200$, learning rate $\mu = 0.8$, moment term $\alpha = 0.7$, and the error tolerances are (1) $\tau = 1e - 5$; (2) $\tau = 0$ respectively.

The corresponding outputs of two experiments are as shown in the last two rows of Table 1.

| Inputs | Desired | Output1 | Output2 |
|--------|---------|---------|---------|
| 0.1    | 0.1     | 0.1013  | 0.1000  |
| 0.1    | 0.95    | 0.9468  | 0.9500  |
| 0.95   | 0.1     | 0.9468  | 0.9500  |
| 0.95   | 0.95    | 0.1048  | 0.1000  |

Fig. 3 shows the training errors of the network output and desired output in the supervised algorithm. The test input samples and output as shown in Table 2.

| Inputs | Desired | Output1 | Output2 |
|--------|---------|---------|---------|
| 0.95   | 0.95    | 0.1048  | 0.1000  |
| 0.1    | 0.1     | 0.1013  | 0.1000  |
| 0.1    | 0.95    | 0.9468  | 0.9500  |
| 0.95   | 0.1     | 0.9468  | 0.9500  |

The average result of ten folds for SANN to solve the XOR problem with architecture 2-1-1 of $\tau = 0$ are as follows: running time is 0.0812 s, training mse is 1.77919e-25, test mse is 1.89882e-25.
A Novel Spatial Architecture ANNs Based on Multilayer Feedforward Network

From Table 1-2, and Fig. 3 we can find that a SANN network with only one hidden unit can solve the xor problem at any training accuracy, meanwhile, the similar test accuracy was obtained. Besides, the convergence is very fast that it only needs several iterations.

(2) 2-2-1 BPN and SANN
The smallest architecture of multilayer feedforward network is with two hidden units. Here, we compare the generalization capability of SANN and BPN with the architecture of 2-2-1.

Set the value of global parameters are as follows: max iteration number $T = 2000$, learning rate $\mu = 0.8$, moment term $\alpha = 0.7$, and the error tolerances $\tau = 0$. The training mse is shown in Fig. 4, and the average result of ten folds for SANN and BPN respectively is given by in Table 3.

|                     | BP Network | SANN       |
|---------------------|------------|------------|
| Global Minimum      | 3.2741e-33 | 1.10741e-33|
| Iteration Number    | 418        | 603        |
| Training MSE        | 3.67811e-31| 3.53171e-32|
| Test MSE            | 6.5482e-32 | 4.2695e-32 |

From this experiment we can find that when they both have two hidden neurons to solve the xor problem, the SANN network has the similar training and test accuracy with BP network, and the convergence of SANN network is faster than BP network, but it need more training time because there are more weights should be calculated than BP network in the learning.
Fig. 4. The training mse of SANN network (2-2-1) for XOR problem. The architecture is two input neurons, one hidden neurons and one output neurons. $T = 200$, $\mu = 0.8$, $\alpha = 0.7$, $\tau = 0$.

Fig. 5. The training mse of SANN and BP network (8-4-1) with the parameters of $T = 1000$, $\tau = 1e-05$

5.2 Approximation Capability

Consider the nonlinear function: $f(x) = \frac{x_1 x_2 + x_3 x_4 + x_5 x_6 + x_7 x_8}{400}$. Where $x_i (i = 1, 2, ..., 8)$ is drawn at random from $[0, 10]$. Similar mapping functions have been used in [16-18]. In this paper we use 250 samples to train a SANN network of size 8-4-1 (same with [17]). The parameters are set as: the learning error tolerance $\tau = 1e - 5$, max iteration number $T = 1000$, learning rate $\mu = 1e - 4$, moment term $\alpha = 0.7$. The training data set are sampled uniformly in the domain. In order to estimate the generalization capability of SANN network, an independent set of 500 test data is generated at random. The weights are initialized as the
training mse = 9.0026e-04, test mse = 0.0019

**Fig. 6.** The training mse of SANN (8-2-1) with the parameters of $T = 1000, \tau = 1e-05$ random numbers distributed in the interval [-1,1]. As a comparison, a threelayer feed-forward neural network is used and trained using BP algorithm with momentum.

Table 4 lists the comparison results of the SANN network and BP network. The training error results are shown in Fig. 5.

Besides, we also use a SANN network with two hidden units to approximate the nonlinear function, the training error is shown in Fig. 6.

In this experiment, we can find that for a given training accuracy and finite iteration number, SANN network need little hidden units than multilayer feed-forward network. Whether has the same number of hidden units or less than, SANN network performs better than the multilayer feedforward network in terms of both the approximation capability and generalization capability. Besides, SANN can alleviate or avoid the local minima problem.

### 6 Conclusion and Discussion

In this paper, a novel topology architecture of artificial neural network is proposed based on multilayer feedforward network, lateral inhibition mechanism and spatial span connection mode, and named SANN. The activation function of input neurons is tunable and according to the priori knowledge to chose a proper function as the input neurons’ activation function. In this paper, when they connect to the output neurons, choose a nonlinear function, otherwise, choose a linear function. The proposed network shown higher representation and generalization capability in the numerical experiments, and can alleviate or avoid the local minima problem than multilayer feedforward network. But because of the more synapse connections than feedforward network, it needs more time to learn the samples. Design a suitable learning algorithm for SANN is the next useful work.
Acknowledgments. This study is supported by the NSFC (60873043), 863 programs of China (2007AA04Z160), BMSF (4092010), DFME (200800050004), BMET, NCET.

References

1. McCulloch, W., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5, 115–433 (1943)
2. Anders, K.: What are Artificial Neural Networks? Nature Biotechnology 26, 195–197 (2008)
3. Stamatis, N., Parthimos, D., Griffith, T.M.: Forecasting chaotic cardiovascular time series with an adaptive slope multilayer perceptron neural network. IEEE Transactions on Biomedical Engineering 46(12), 144–153 (1999)
4. Li, Z., Melek, W.W.: Neural network-based prediction of cardiovascular response due to the gravitational effects. Control and Intelligent Systems 36(1), 81–91 (2008)
5. Liu, D., Ju, C.: Application of an improved BP neural network in business forecasting. In: Sixth World Congress on Intelligent Control and Automation, p. 5 (2006)
6. Pan, W.-T., Lin, W.-Y.: Use probabilistic neural network to construct early warning model for business financial distress. In: 2008 International Conference on Management Science and Engineering (ICMSE), pp. 134–139 (2008)
7. Pan, W.-T.: Use of probabilistic neural network to construct early warning model for business financial distress. Journal of Statistics and Management Systems 11(4), 749–760 (2008)
8. Cheng, M.-Y., Tsai, H.-C., Sudjono, E.: Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. Expert Systems with Applications 37(6), 4224–4231 (2010)
9. Chao, J., Hoshino, M., Kitamura, T.: A New Pyramid Network and Its Generalization Performance. In: Proc. of IJCNN 2001: International Joint Conference on Neural Networks, Washington D.C., pp. 2811–2816 (2001)
10. Jaeger, H., Haas, H.: Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. Science 304, 78–80 (2004)
11. Buzsáki, G.: The structure of consciousness. Nature 446, 267 (2007)
12. Sa, L., Kil, R.M.: A Gaussian potential function network with hierarchically self-organization learning. Neural Networks 4, 207–224 (1991)
13. Simon, W.E., Carter, J.R.: Back propagation learning equations from the minimization of the recursive error. In: IEEE International Conference on Systems Engineering, pp. 155–160 (1989)
14. Jain, L.C.: Hybrid Connectionist Systems in Research and Teaching. IEEE Aerospace and Electronic Systems Magazine 10(3), 14–18 (1995)
15. Li, A.-J., Liu, Y.-H., Luo, S.-W.: On the solution of the XOR problem using the decision tree-based neural network. In: International Conference on Machine Learning and Cybernetics, vol. 2, pp. 1048–1052 (2003)
16. Barmann, F., Biegler-Konig, F.: On a class of efficient learning algorithms for neural networks. Neural Networks 5, 139–144 (1992)
17. Gad, E.F., Atiya, A.F., Shaheen, S., El-Dessouki, A.: A new algorithm for learning in piecewise-linear neural networks. Neural Networks 13, 485–505 (2000)
18. Wen, C., Ma, X.: A max-piecewise-linear neural network for function approximation. Neurocomputing 71(4-6), 843–852 (2008)