Article

Invertebrate diversity classification using self-organizing map neural network: with some special topological functions

WenJun Zhang[1,2], **QuHuan Li**[1]

[1]School of Life Sciences, Sun Yat-sen University, Guangzhou, China; [2]International Academy of Ecology and Environmental Sciences, Hong Kong

E-mail: wjzhang@iaees.org,zhwj@mail.sysu.edu.cn

Received 16 March 2014; *Accepted 10 April 2014*; *Published online 1 June 2014*

Abstract

In present study we used self-organizing map (SOM) neural network to conduct the non-supervisory clustering of invertebrate orders in rice field. Four topological functions, i.e., cossintopf, sincostopf, acossintopf, and expsintopf, established on the template in toolbox of Matlab, were used in SOM neural network learning. Results showed that clusters were different when using different topological functions because different topological functions will generate different spatial structure of neurons in neural network. We may chose these functions and results based on comparison with the practical situation.

Keywords self-organizing map (SOM) neural network; topological functions; Matlab; cluster analysis; invertebrates.

Selforganizology
URL: http://www.iaees.org/publications/journals/selforganizology/online-version.asp
RSS: http://www.iaees.org/publications/journals/ selforganizology /rss.xml
E-mail: selforganizology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Invertebrate diversity in the farmland is always a research focus for its important role in maintaining natural equilibrium (Brown, 1991; Kremen et al., 1993; Way and Heong, 1994; Zhang and Barrion, 2006). We always understand invertebrate diversity by field sampling. However, sampling information is overall non-linear and could not be treated by linear or parametric statistic methods (Maravelias et al., 2003). Artificial intelligence is usually used to learn knowledge from complex information systems (Zhang and Li, 2006; Zhang, 2007a, b; Zhang et al., 2007; Zhang et al., 2008a, b; Zhang and Zhang, 2008; Zhang, 2010, 2011, 2012). They are paralleled and distributed models. Knowledge and information are stored and distributed in the weights of neural networks through learning from samples (Bian and Zhang, 2000). In the artificial neural network, the structure of neurons may be defined in different ways. The topological function is one of the important ways. Topological functions are used to generate spatial and topological structure of neurons. Different choices of various topological functions would result in different results in neural network learning.

In present study we use self-organizing map (SOM) neural network to conduct the non-supervisory clustering of invertebrate orders in rice field. Several topological functions are used in SOM neural network learning (Zhang and Li, 2006). We can choose these functions and results based on the comparison with the practical situation.

2 Materials and Methods

2.1 Sampling data

Invertebrate biodiversity in four periods were investigated in the rice fields of IRRI (International Rice Research Institute). Rice invertebrates were collected by a sampler and 60 samples were surveyed. Invertebrates were identified, sorted, and counted. The raw data were lumped using LUMP (Schoenly and Zhang, 1999) and the data for invertebrate orders (or the equivalent taxa, abbreviated in the following) were obtained. There were totally 21 orders.

2.2 Topological functions

Four topological functions were used, which were established on the template of topological function, mytopf, in Matlabneural network toolkit (Fecit, 2003), as indicated in the following:

cossintopf, major mathematical function: cos(sin(cx));
sincostopf, major mathematical function: sin(cx) + cos(cx);
acossintopf, major mathematical function: acos(sin(cx));
expsintopf, major mathematical function: $e^{\sin(cx)}$.

Their Matlab codes are summarized as below.

```
%Topological function cossintopf
function pos=cossin(varargin)
%Custom topology function.
%   Syntax
%       pos = cossin(dim1,dim2,...,dimN)
%           dimi - number of neurons along the ith layer dimension
%           pos    - NxS matrix of S position vectors, where S is the
%                     total number of neurons which is defined by the
%                       product dim1*dim1*...*dimN.

%Example
%       pos = cossin(20,20);
%       plotsom(pos)
% $Revision: 1.2 $
dim = [varargin{:}];       % The dimensions as a row vector
size = prod(dim);           % Total number of neurons
dims = length(dim);         % Number of dimensions
pos = zeros(dims,size); % The size that POS will need to be set
len = 1;
pos(1,1) = 0;
for i=1:length(dim)
   dimi = dim(i);
   newlen = len*dimi;
   pos(1:(i-1),1:newlen) = pos(1:(i-1),rem(0:(newlen-1),len)+1);
```

```
    posi = 0:(dimi-1);
    pos(i,1:newlen) = posi(floor((0:(newlen-1))/len)+1);
    len = newlen;
end
for i=1:length(dim)
    pos(i,:)=pos(i,:)*0.7+cos(sin([1:size]*exp(1)/5*i))*0.3;
end
```

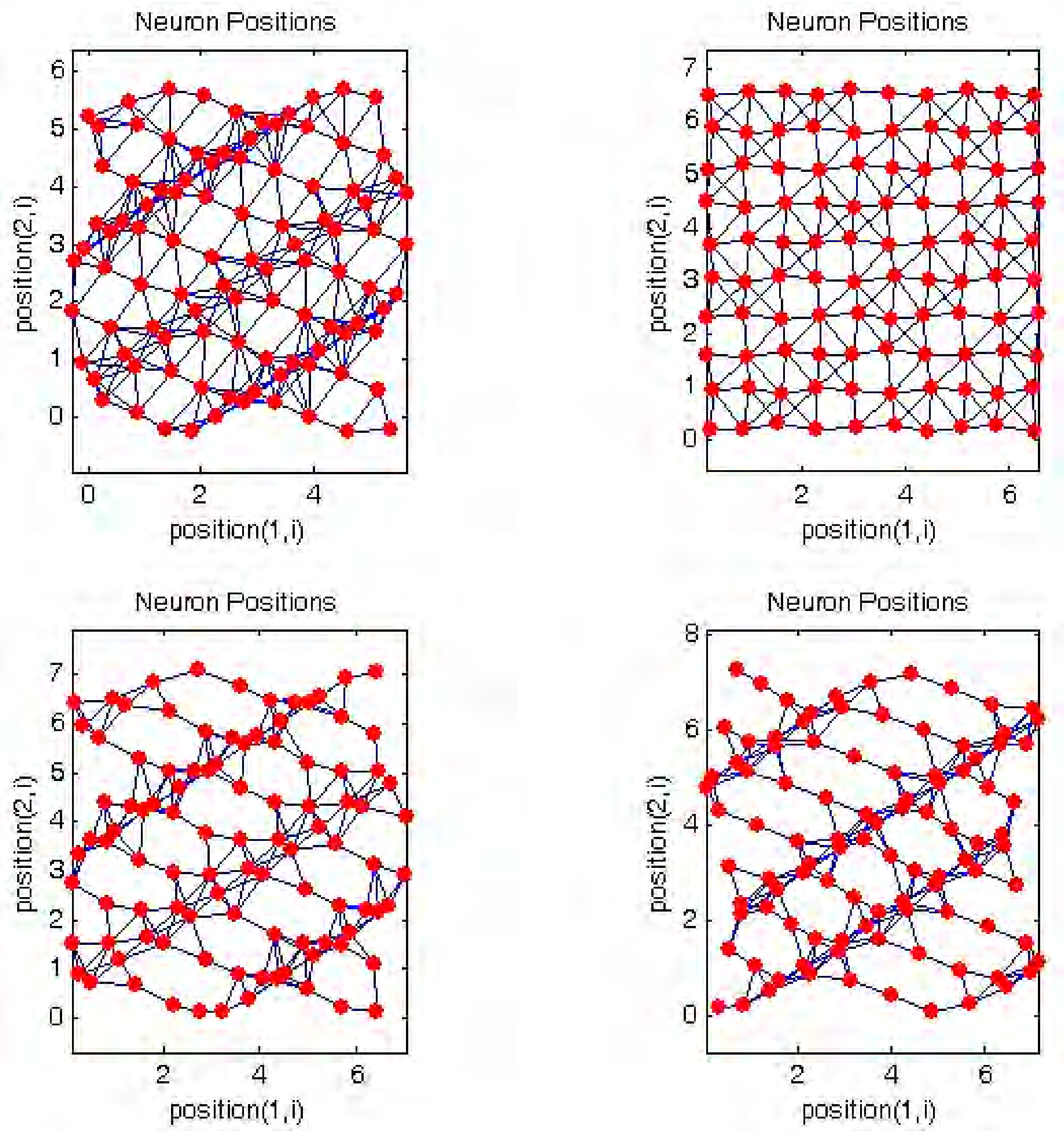

Fig. 1 Two-dimensional topological structures of neurons, generated from various topological functions. Upper left: sincostopf; upper right: cossintopf; lower left: expsintopf; Lower right: acossintopf.

2.3 Source codes

```
%Topological function sincostopf
%The source codes are the same as in 2.2
for i=1:length(dim)
pos(i,:)=pos(i,:)*0.6+sin([1:size]*exp(1)/5*i)*0.2+cos([1:size]*exp(1)/5*i)*0.2;
end

%Topological function acossintopf
% The source codes are the same as in 2.2
for i=1:length(dim)
   pos(i,:)=pos(i,:)*0.7+acos(sin([1:size]*exp(1)/5*i))*0.3;
end

%Topological function expsintopf
% The source codes are the same as in 2.2
for i=1:length(dim)
   pos(i,:)=pos(i,:)*0.7+exp(sin([1:size]*exp(1)/5*i))*0.3;
end
```

Topological structures generated from four topological functions are different, as illustrated in Fig. 1.Various topological functions demonstrated their major differences in types of neuron connection, resulted from various mathematical functions in topological structures. For example, in the two-dimensional topological structures of Fig. 1, topological functions cossintopf and and sincostopf are significantly different from the other two functions, i.e., each neuron has not less than two connections for cossintopf and sincostopf, however, in other two functions there are neurons with only one connection. Both the number of neuron connections and connected neurons for the given neuron are different among four topological functions.

2.4 Self-organizing map neural network

The four topological functions were used to generate topological structures of neurons in self-organizing map (SOM) neural network, which can be used for the unsupervised self-organization study of two-dimensional SOM. Matlab codes are listed as follows:

```
P = MarOIDs(:,:);
P = P';
%Generate neural network
net = newsom(minmax(P),[8 8]); %Set 8*8 neurons
net.layers{1}.topologyFcn = 'cossintopf'; %Set cossintopf as the topological function
%Obtain connectivity weights
winit = net.iw{1,1};
%Train neural network
net.trainParam.epochs = 1000;    %Set training epochs=1000
net = init(net);
net = train(net,P);
%Obtain connectivity weights trained
w = net.iw{1,1}
%Input sample vectorand updated weights
```

```
cluster = 0;
for i = 1:size(P,2);
cluster = vec2ind(sim(net,P(:,i)));
outputclass(1,i) = i;
outputclass(2,i) = cluster;
end
outputclass
```

3 Results

Different from the classification of functional groups in which the supervised clustering may be used, there are diverse and complex relationships between different sub-classification taxa in the same order (Schoenly and Zhang, 1999). For example, there are predators, preys, and competitors in Coleoptera. Therefore, it is reasonable to use unsupervised clustering at the order level, in order to reflect between-order relationships.

Table 1 Self-organizing cluster analysis of rice invertebrates using various topological functions.

	sinco stopf				cossi ntopf				acossi ntopf				exps intopf				Default			
Order	Mar	Apr	Sep	Oct	Mar	Apr	Sep	Oct	Mar	Apr	Sep	Oct	Mar	Apr	Sep	Oct	Mar	Apr	Sep	Oct
Lepidoptera	5	10	9	34	17	57	48	18	51	7	8	3	53	59	62	43	47	49	10	55
Ephemeroptera	4	19	10	17	41	33	38	9	57	11	20	1	59	35	55	57	32	59	33	64
Hemiptera	64	64	64	64	64	8	1	58	1	63	57	57	9	16	5	16	1	64	64	1
Orthoptera	4	1	34	8	25	43	54	31	57	5	48	21	59	53	45	61	23	50	13	60
Hymenoptera	10	36	4	28	1	37	63	34	41	26	32	10	64	38	59	36	63	38	20	54
Diptera	24	24	8	46	45	64	58	62	7	33	56	33	22	41	27	40	29	5	7	20
Odonata	3	34	4	1	35	25	63	25	59	20	40	1	61	45	35	58	46	53	20	63
Coleoptera	26	39	48	30	5	40	49	64	37	49	43	35	39	27	21	21	53	32	23	29
Araneae	44	48	36	16	12	30	52	38	27	57	26	25	24	21	40	27	43	16	45	44
Dermaptera				33				9				1				57				64
Strepsiptera			18	33			31	9			8	1			64	57			1	64
Acari	17	16	2	11	3	61	64	20	33	32	16	13	48	48	53	51	62	28	4	39
Neuroptera			18				31				8				64				1	
Thysanoptera	4	10		17	25	57		1	57	7		1	59	59		57	23	49		64
Uniden. order	48	17	24	32	31	41	33	8	18	13	63	8	27	51	17	48	20	57	16	8
Isoptera			17	9			24	1			7	1			63	57			9	64
Mesogastropoda	16	57	57	57	8	10	8	49	64	1	10	56	41	1	41	1	8	1	57	57
Arthropleona	2	4	1	4	9	59	14	5	50	8	5	24	54	62	48	59	56	35	17	32
Blattodea	4		27	18	25		31	3	57		8	2	59		64	41	23		1	48
Cyproida	57	10	3		57	57	28		32	7	30		16	59	59		57	49	35	
Siphonaptera	21	3			18	49			61	15			51	64			48	33		

"Default" means default setting of topological function in Matlab. The values mean the response neurons, i.e., the categories which orders belong to.

Using the SOM with different topological functions as described above, and with other default functions in the SOM of Matlab, a self-organizing unsupervised clustering was conducted on invertebrate orders based on the aforementioned source codes. The results for the neural networks with four topological functions and default functions were obtained (Table 1), and summarized as follows:

(1) March

Using topological function sincostopf: (Ephemeroptera, Orthoptera, Thysanoptera, Blattodea), the rest of the orders were of the same category;

Using topological function cossintopf: (Ephemeroptera, Orthoptera, Thysanoptera, Blattodea), the rest of the orders were of the same category;

Using topological function acossintopf: (Ephemeroptera, Orthoptera, Thysanoptera, Blattodea), the rest of the orders were of the same category;

Using topological function expsintopf: (Ephemeroptera, Orthoptera, Thysanoptera, Blattodea), the rest of the orders were of the same category;

System default function: (Orthoptera, Thysanoptera, Blattodea), the rest of the orders were of the same category.

(2) April

Using topological function sincostopf: (Lepidoptera, Thysanoptera, undetermined order), the rest of the orders were of the same category;

Using topological function cossintopf: (Lepidoptera, Thysanoptera, undetermined order), the rest of the orders were of the same category;

Using topological function acossintopf: (Lepidoptera, Thysanoptera, undetermined order), the rest of the orders were of the same category;

Using topological function expsintopf: (Lepidoptera, Thysanoptera, undetermined order), the rest of the orders were of the same category;

System default function: (Lepidoptera, Thysanoptera, undetermined order), the rest of the orders were of the same category.

(3) September

Using topological function sincostopf: (Hymenoptera, Odonata), (Strepsiptera, Neuroptera), the rest of the orders were of the same category;

Using topological function cossintopf: (Hymenoptera, Odonata), (Strepsiptera, Neuroptera), the rest of the orders were of the same category;

Using topological function acossintopf: (Lepidoptera, Strepsiptera, Neuroptera, Blattodea), the rest of the orders were of the same category;

Using topological function expsintopf: (Hymenoptera, undetermined order), (Strepsiptera, Neuroptera, Blattodea), the rest of the orders were of the same category;

System default: (Hymenoptera, Odonata), (Strepsiptera, Neuroptera, Blattodea), the rest of the orders were of the same category.

(4) October

Using topological function sincostopf: (Ephemeroptera, Thysanoptera), (Dermaptera, Strepsiptera), the rest of the orders were of the same category;

Using topological function cossintopf: (Ephemeroptera, Dermaptera, Strepsiptera), the rest of the orders were of the same category;

Using topological function acossintopf: (Ephemeroptera, Odonata, Dermaptera, Strepsiptera, Thysanoptera, Blattaria), the rest of the orders were of the same category;

Using topological function expsintopf: (Ephemeroptera, Dermaptera, Strepsiptera, Thysanoptera, Blattaria), the rest of the orders were of the same category;

System default function: (Ephemeroptera, Dermaptera, Strepsiptera, Thysanoptera, Blattaria), the rest of the orders were of the same category.

It can be founded that the general trends from various classifications are similar to each other. However, the results between four topological functions and between four topological functions and the default function are somewhat different.

4 Conclusions and Discussion

In the SOM, different topological functions can be set and unsupervised self-organizing clustering may thus be conducted. The same results may be used but the classification differences should be analyzed for further analysis. The biological meanings for these differences may be explored to find some principles and mechanisms. It is suggested that two principles, i.e., topological connectivity and topological symmetry of topological structure (Lin, 1998) in the neural network, should be followed in the establishment of topological functions.

References

Bian JQ, Zhang XG. 2000. Pattern Recognition (Second edition). Tsinghua University Press, Beijing, China

Brown KS Jr. 1991. Conservation of neotropical insects: Insects as indicators. In: The Conservation of Insects and Their Habitats (Collins NM, Thomas, JA, eds). 349-404, Academic, London, UK

Feci. 2003. MATLAB6.5 Auxiliary Neural Network Analysis and Design. 165-187, Electronic Industry Publishing House, Beijing, China

Kremen C, Colwell RK, Erwin TL, et al. 1993. Invertebrate assemblages: Their use as indicators in conservation planning. Conservation Biology, 7: 796-808

Lin JK. 1998. Elementary Topology. Science Press, Beijing, China

Maravelias CD, Haralabous J, Papaconstantinou C. 2003. Predicting demersal fish species distributions in the Mediterranean Sea using artificial neural networks. Marine Ecology Progress series, 255: 249-258

Schoenly K G., Zhang W J. IRRI Biodiversity Software Series. I. LUMP, LINK, AND JOIN: utility programs for biodiversity research. IRRI Technical Bulletin No. 1. Manila (Philippines): International Rice Research Institute, 1999.1-23.

Way MJ, & Heong KL. 1994. The role of biodiversity in the dynamics and management of insect pests of tropical irrigated rice - A review. Bulletin of Entomological Research, 84: 567-587

Zhang WJ. 2007a. Pattern classification and recognition of invertebrate functional groups using self-organizing neural networks. Environmental Monitoring and Assessment, 130:415-422

Zhang WJ. 2007b. Supervised neural network recognition of habitat zones of rice invertebrates. Stochastic Environmental Research and Risk Assessment, 21: 729-735

Zhang WJ. Computational Ecology: Artificial Neural Networks and Their Applications. World Scientific, Singapore, 2010

Zhang WJ. 2011. Simulation of arthropod abundance from plant composition. Computational Ecology and Software, 1(1):37-48

Zhang WJ. 2012. Computational Ecology: Graphs, Networks and Agent-based Modeling. World Scientific, Singapore, 2012

Zhang WJ, Bai CJ, Liu GD. 2007. Neural network modeling of ecosystems: a case study on cabbage growth system. Ecological Modelling, 201:317-325

Zhang WJ, Li QH. 2006. Development of topological functions in neural networks and their application in SOM learning to biodiversity. Computer Applications and Software, 23(10): 71-73

Zhang WJ, Liu GH, Dai HQ. 2008a. Simulation of food intake dynamics of holometabolous insect using functional link artificial neural network. Stochastic Environmental Research and Risk Assessment, 22(1): 123-133

Zhang WJ, Zhang XY. 2008. Neural network modeling of survival dynamics of holometabolous insects: a case study. Ecological Modelling, 211,433-443

Zhang WJ, Zhong XQ, Liu GH. 2008b. Recognizing spatial distribution patterns of grassland insects: neural network approaches. Stochastic Environmental Research and Risk Assessment, 22(2): 207-216