Vol. 47 No. SIG 1(TOM 14) IPSJ Transactions on Mathematical Modeling and Its Applications Feb. 2006

Regular Paper

Spherical SOM and Arrangement of Neurons Using Helix on Sphere

Hirokazu Nishio,† Md. Altaf-Ul-Amin,† Ken Kurokawa†
and Shigehiko Kanaya†

Self-organizing maps (SOM) are a kind of neural network, and are applied to many fields. In many applications, there are borders surrounding the neuron arrangement. It causes a problem which is called 'Border effect' because the number of neighborhood neurons of a neuron near a border is different from that of a neuron near the center. This problem can be solved by arranging neurons uniformly on the surface of a sphere. But by the conventional method we cannot arrange neurons of arbitrary number. Therefore, it is inconvenient to use. Here we developed a method to arrange the neurons of arbitrary number by dividing a helix which covers the surface of a sphere into equal parts. It also can arrange neurons on a sphere more uniformly than the conventional method.

1. Introduction

Self-organizing maps (SOM) are a kind of neural network, and are applied to many fields for understanding the distribution in high dimensional space. In many applications, its neurons are arranged on plain surface and there are borders surrounding the neuron arrangement. It causes a problem which is called 'Border effect' because the number of neighborhood neurons of a neuron near a border is different from that of a neuron near the center [1]. Though this problem can be solved by arranging neurons uniformly on the surface of a sphere [2], a disadvantage of the conventional method is that it cannot arrange neurons of arbitrary number. In this paper we propose a method that can arrange neurons of arbitrary number. In this method we arrange the neurons by dividing a helix that covers the surface of sphere into equal parts.

2. Conventional Method

In case of spherical SOM proposed by Ritter, the neurons are arranged by subdividing an icosahedron recursively as shown in **Fig. 1** [2]. This method is commonly used for the application of spherical SOM [3),4]. We call the arrangement by this method as $ICOSA_N$, where N is the number of recursive subdivision. $ICOSA_N$ can arrange $2+10\cdot 4^N$ neurons. **Table 1** shows the number of neurons in $ICOSA_N$ for some typical values of N. Since it increases exponentially, sometimes we can't obtain a suitable arrangement for useful number of neurons.

3. Proposed Method

First, we consider a helix which goes around a sphere of unity radius (See **Fig. 2** (a)). The helix is determined by the equation below.

$$\theta = 2k\phi (0 \le \phi \le \pi)$$

In the above equation, k is the number of turns in the helix and spherical coordinates θ and ϕ can be transformed to orthogonal coordinates using the following equations (See Fig. 2 (b)).

$$x = \cos\theta \sin\phi,$$
$$y = \sin\theta \sin\phi,$$
$$z = \cos\phi$$

There is no restriction for the value of k. In this paper we set k as $\sqrt{n}$ where n is the number of neurons. The reason of this will be discussed later. Second, we calculate the helix length L. Since it is difficult to calculate it analytically, we calculate it by numerical integration. Finally, we arrange neurons at equal intervals $L/(n-1)$ on this helix. It is easy to realize that any number of neurons can be suitably arranged on the map by following the proposed method. Let us now consider why we set k as $\sqrt{n}$. **Figure 3** shows the relation between the length and the number of turns of the proposed helix. The helix length L is mostly proportional to k in practical range of k. And it can be shown that the helix pitch (see **Fig. 4** (a)) is mostly inversely proportional to k. So when we set k as $\sqrt{n}$, the ratio of neuron interval (NI) and helix pitch (HP) is preserved for any value

† Nara Institute of Science and Technology, Graduate School of Information Science

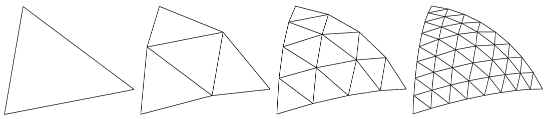
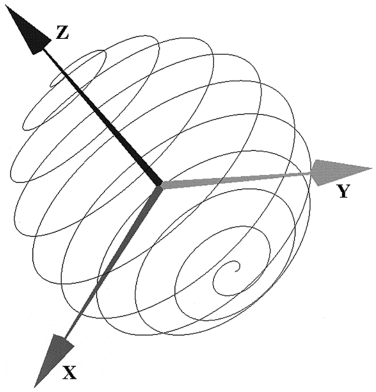


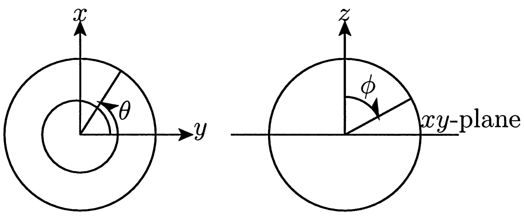
Fig. 1 Recursive subdivision of a triangle of an icosahedron.

Table 1 Relation between N and the number of neurons in $ICOSA_N$.

N	number of neurons
0	12
1	42
2	162
3	642
4	2,562
5	10,242



(a) A helix on a sphere



(b) Relation between axis x, y, z, θ, ϕ

Fig. 2 A helix on a sphere and its coordinates.

of n. $NI = \frac{L}{n-1} \propto \frac{K}{n-1} = \frac{\sqrt{n}}{n-1} \approx \frac{1}{\sqrt{n}}$ and, $HP \propto \frac{1}{k} = \frac{1}{\sqrt{n}}$ therefore, $\frac{NI}{HP} \propto \frac{1}{\sqrt{n}} \times \frac{\sqrt{n}}{1} = 1$ i.e., $\frac{NI}{HP}$ is constant. Figures 4 (b) and 4 (c) shows the arrangements of neurons for $k = \sqrt{n}$ and Fig. 4 (d) shows the same for $k \neq \sqrt{n}$. The advantage of setting k as $\sqrt{n}$ is that the shape of the dots representing the neurons on the final map remains unchanged for different values of n. This makes the visualization quality of the map better.

4. Uniformity of Neuron Arrangement

In this section we will compare the conventional method with the proposed method in the context of uniformity of the arrangements of the neurons. We use $HELIX_N$ to mean the arrangement by proposed method which has the same number of neurons as $ICOSA_N$ has. Uniformity means the equality of the number of neighborhood neurons in the context of all the neurons in the map. The map is completely uniform in the case when the number of neighborhood neurons is the same for each of the neurons of the map. Let $f(\eta, r)$ is the numbers of neurons which are within radius r from neuron η (See **Fig. 5**). And let $V(r)$ is the variance of $f(\eta, r)$ for all neurons, i.e.,

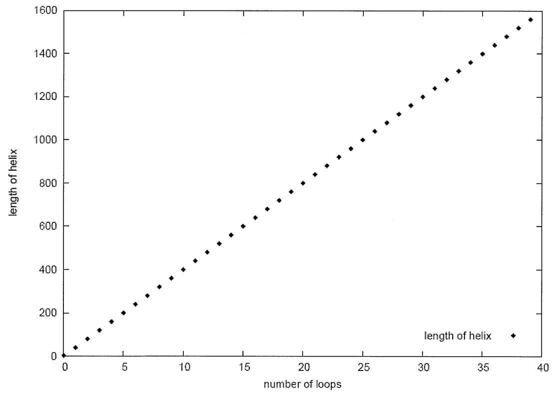


Fig. 3 Relation between nunber of loops k and the helix length L.

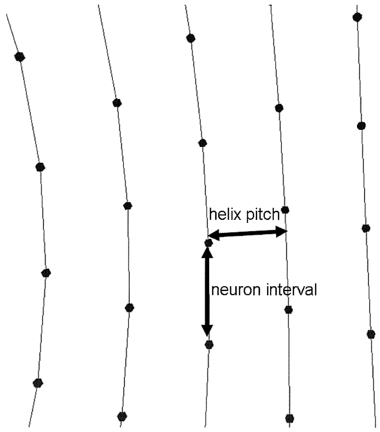
$$V(r) = \sum_{\eta} \frac{(f(\eta, r) - \overline{f(\eta, r)})^2}{n}$$

The variance $V(r)$ can only be a real positive number or zero. When the number of neighborhood neurons $f(\eta, r)$ is the same for all neuron η, $V(r)$ is zero. The relations between the radius r and the variance $V(r)$ are shown in **Fig. 6** (a). It shows $V(r)$ in the case of helix is mostly lower than that in the case of icosahedron.

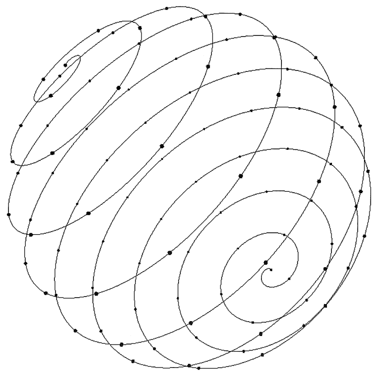
We propose *Untidiness* as a measure of uniformity and define *Untidiness* as follows:

$$Untidiness := \int_{\theta=0}^{\theta=\pi/2} V\left(2\sin\left(\theta/2\right)\right) \mathrm{d}\theta$$

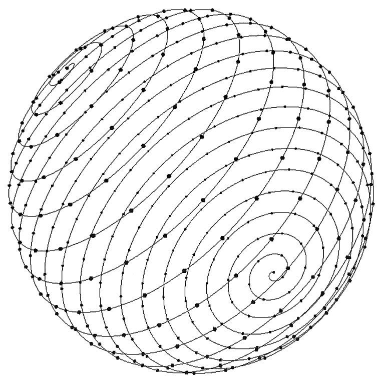
Here $2\sin(\theta/2)$ is the length of chord for center angle θ. *Untidiness* can only be a real positive number or zero. The more uniform is the neuron arrangement the less is the value of untidiness. When the arrangement is completely uniform *Untidiness* is zero. The following **Table 2** is the *Untidiness* of $ICOSA_N$ and $HELIX_N$ for N in range $0 \leq N \leq 5$.



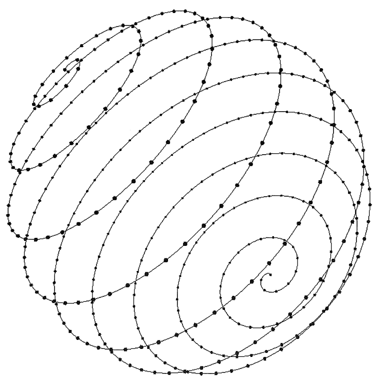
(a) Helix pitch and neuron interval



(b) Arrangement when $n = 100$ and $k = 10$



(c) Arrangement when $n = 400$ and $k = 20$



(d) Arrangement when $n = 400$ but $k = 10$

Fig. 4 Relation between arrangement and parameters.

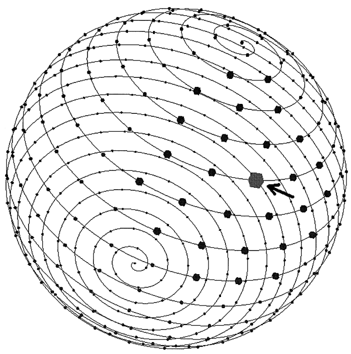


Fig. 5 28 neurons are within radius 0.5 from the indicated neuron η, i.e., $f(\eta, 0.5) = 28$.

$ICOSA_0$ is completely uniform because it is icosahedron itself. And $ICOSA_1$ has smaller *Untidiness* than $HELIX_1$. Since *Untidiness* of $ICOSA_N$ increases quickly, it is larger than that of $HELIX_N$ when N is two or more. **Figure 7** shows the relation between the number of neurons and *Untidiness*. By conventional method we can't arrange arbitrary number of neurons.

Table 2 *Untidiness* of $ICOSA_N$ and $HELIX_N$ for $0 \le N \le 5$.

N	*Untidiness* of $ICOSA_N$	*Untidiness* of $HELIX_N$
0	0.0000	0.3611
1	0.4253	0.4539
2	2.8707	1.3165
3	3.9360	2.3874
4	10.4419	4.4689
5	82.1324	9.5022

So the *Untidiness* for the permissible number of neurons are shown by crosses on the dotted line. On the other hand, we can arrange any number of neurons by proposed method and the solid line shows the relation between the number of neurons and untidiness. The *Untidiness* in case of the proposed method is much lower compared to the conventional method, indicating that the proposed method has an advantage in uniformity of arrangement. It is considered that the regularity of arrangement may be important in view of visibility, and conventional method might be more regular. The variance of $ICOSA_N$ is less than that of $HELIX_N$ on

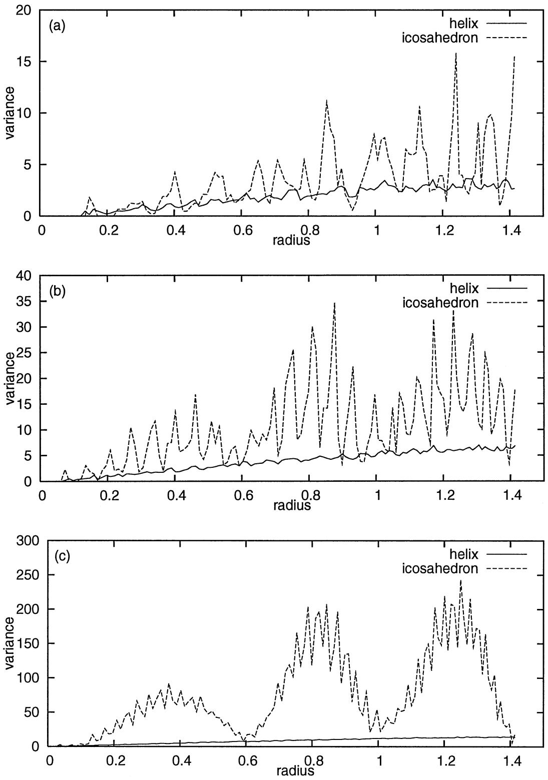


Fig. 6 Relation between the radius and the variance. (a) $ICOSA_3$ and $HELIX_3$, (b) $ICOSA_4$ and $HELIX_4$, (c) $ICOSA_5$ and $HELIX_5$.

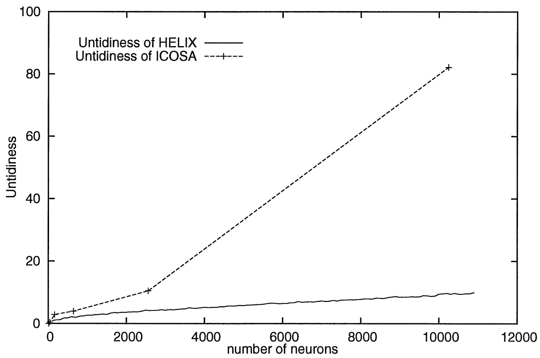


Fig. 7 Relation between *Untidiness* and the number of neurons.

some particular radii. It means that both regularity and uniformity are satisfied when we use those radii only. The SOM which use conventional method can avoid border effect by using those radius for learning process.

5. Conclusion

In this paper we propose a new approach for the arrangement of neurons concerning spherical SOM. According to the proposed approach the neuron are placed at equal distance on a helix that goes around a sphere. This enables us to suitably arrange an arbitrary number of neurons which is not possible in conventional icosahedron method. Another advantage of the proposed approach is the uniformity of the arrangement. The arrangement by the proposed approach is more uniform than the icosahedron approach when the number of neuron are 43 or more. Thus the proposed approach relaxes the usability and improves the usefulness of the spherical SOM.

References

1) Kohonen, T.: *Self-Organizing Maps*, Springer Series in Information Sciences, 2nd ed., Springer (1997).
2) Ritter, H.: *Self-Organizing Maps on non-euclidean Spaces*, Kohonen Maps, pp.97–108, Oja, E. and Kaski, S. (eds.) (1999).
3) Nakatsuka, D. and Oyabu, M.: Application of Spherical SOM in Clustering, *Proc. WSOM'03*, pp.203–207 (2003).
4) Nakatsuka, D. and Tokutaka, H.: Visualization and Cluster Analysis using Spherical SOM, *Genome Informatics* (2004).
5) Nishio, H., Amin, M., Sato, T., Wada, K., Wada, Y., Minato, K., Kobayashi, K., Ogasawara, N. and Kanaya, S.: Visualization of gene classification based on expression profile using BL-SOM, *Proc. WSOM03*, 101–106 (2003).
6) Nishio, H., Wada, K., Wada, Y., Amin, M. and Kanaya, S.: Suitability of Spherical SOM for Gene Expression Analysis, *Proc. RECOMB2004* (2004).
7) Nishio, H., Altaf-Ul-Amin, M., Nakamura, Y., Abe, T., Kinouchi, M., Ikemura, T., Kobayashi, K., Ogasawara, N. and Kanaya, S.: Gene classification Based on Expression Profile using BL-SOM: Suitability assessment of multivariate gene expression data to Spherical and Plain SOM by N-measure, *Proc. SCI2004* (2004).

(Received December 13, 2004)
(Revised May 12, 2005)
(Accepted June 8, 2005)

Hirokazu Nishio was born in 1981. He received B.Sc. degree from Nara Institute of Science and Technology (NAIST) in 2003. From 2004, he has been a Doctoral student at Department of Bioinformatics, Graduate School of Information Science, NAIST. And from 2005, he has been a Reserch Fellow of Japan Society for the Promotion of Science. His research interests include Visualization of Large-scale data obtained by biological experiment. He is student member of Information Processing Society of Japan, Japan Society of Bioinfomatics, International Society for Computational Biology and Japan Society of Artificial Intelligence.

Md. Altaf-Ul-Amin received B.Sc. in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET) in 1993, Master of Science in Electrical, Electronic and Systems Engineering from Universiti Kebangsaan Malaysia (UKM) in 1999 and Ph.D. from Nara Institute of Science and Technology (NAIST) in 2003. Presently he is working as an Assistant Professor in Comparative Genomics Lab of NAIST. His research interest includes application of Network theory and algorithms and self organizing mapping to bioinformatics.

Ken Kurokawa received Ph.D. in Environmental Science & Microbiology from Graduate School of Pharmacy, Osaka University in 1995. M.Sc. in Earth Physics from Institute of Geology and Paleontology, Graduate School of Science, Tohoku University in 1993, B.Sc. in Structural Geology and Volcanology from Institute of Geology and Paleontology, Faculty of Science, Tohoku University in 1990. He was Associate Professor in Bioinformatics at Graduate School of Information Science, NAIST from 2004, Research Associate in Bioinformatics, Research Institute of Microbial Diseases, Oaka University 2001–2004. Post-Doctoral Fellow in Bioinformatics Genome Information Research Center, Osaka University 1998–2001. His interests are origins and evolution of genome organization. Particularly interested in the correlation between genome information and biosystem evolution. He seeks out the laws of self-organization in genome, and dynamic modeling of biosystems with network thermodynamics.

Shigehiko Kanaya received B.S. from Department Bioscience, Faculty of Science, Science University of Tokyo in 1985, Ph.D. from Department Material and System Engineering, Toyohashi University of Technology in 1990. In 1990, Assistant Researcher in information Engineering at Yamagata University, with a research program on genome informatics concerning species-specific codon usage on the basis of multivariate analysis, and proteome analysis in Onchoryncus species. In 1996, Guest Associate Professor at National Institute of Genetics. In 1999 Associate Professor, Electronic and Information Engineering, and in 2000 Associate Professor, Applied Biosystem Engineering at Yamagata University. In 2000, Guest Researcher at Bioradical Institute (Yamagata Prefecture). In 2001, Associate Professor at Bioinformatics, Research and Education Center for Genetic Information, in 2002, Associate Professor at Comparative Genomics, Department of Bioinformatics, and in 2004–present, Professor at Comparative Genomics, Department of Bioinformatics, Graduate School of Information Science, Nara Institute of Science and Technology.