

# Analysis of DNA Microarray Data by Using Self-Organizing Maps

Tomoyuki Kato<sup>1</sup>

Kikuo Fujimura<sup>1</sup>

Heizo Tokutaka<sup>1</sup>

fujimura@ele.tottori-u.ac.jp

tokutaka@ele.tottori-u.ac.jp

Yasushi Kawata<sup>2</sup>

Masaaki Ohkita<sup>1</sup>

kawata@bio.tottori-u.ac.jp

mohkita@ele.tottori-u.ac.jp

<sup>1</sup> Department of Electrical and Electronic Engineering, Tottori University

<sup>2</sup> Department of Biotechnology and Graduate School of Medical Science, Tottori University

**Keywords:** Self-Organizing Maps, DNA microarray, genes

## 1 Introduction

Self-Organizing Maps (SOM) is a kind of an artificial neural network. SOM is an unsupervised learning. By that, various, multi-dimensional data can be made visible and then clustered. Moreover, by making the component maps, it will be known what kind of influence any arbitrary dimensional component can give for clustering.

Then, an attractively new analytical technique which uses SOM is proposed and used for DNA microarray data in this article. MATLAB was used for the analysis.

## 2 Method and Results

We used datas which analyzed rat RNA samples with DNA microarray of rat 3824 genes. The kinds of genes of data used at this time is 3824 kinds, and the number of samples is 35.

Moreover, the gene name of this sample data was assigned as ID from A1 to A3824. The samples can be clustered according to the following procedures.

1. Each item of 3824 which corresponds to the kind of the gene is assumed to be each dimension of the data vector.
2. The 35 teacher labels according to the samples are located.
  - Thus, the number of data is 35, and each data has a multi-dimensional vector of 3824 dimensions.
3. The result by SOM learning is shown in Fig. 1.

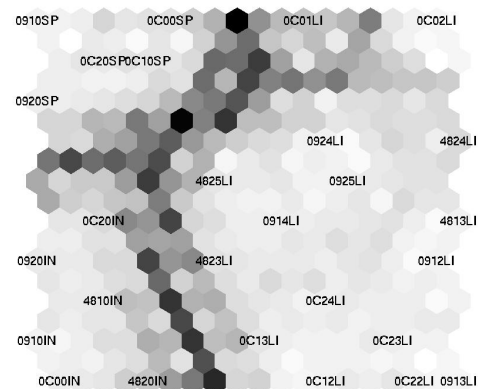


Figure 1: Clustering of samples.

A more accurate phase relation is examined by using spherical SOM in adding to an ordinary plane SOM. The results by a spherical SOM are not presented here. Using this, it was possible to classify 35 conditions results into six classes. These classes are arranged Class A to Class F, as shown in the Table 1, using the result of Fig. 1.

Table 1: Clustering results.

ClassA	ClassB	ClassC	ClassD			ClassE		ClassF	
LI0C01	LI0C02	LI0924	LI4813	LI0913	LI0C14	IN4820	IN0C20	SP0C00	SP4810
	LI0C03	LI0914	LI0912	LI0923	LI4815	IN0C10	IN0920	SP0C10	SP4820
	LI4824	LI4825	LI0C13	LI0C24		IN0910		SP0910	
		LI0925	LI0C12	LI0C22		IN4810		SP0C20	
		LI4823	LI0C23	LI4811		IN0C00		SP0920	

To exclude needless genes for clustering, the following procedures are carried out for the pre-processes.

1. The 35 samples are used as variate, and the variance of each gene is examined.
2. genes which have small standard deviations are excluded. 424 genes which have large standard deviations are chosen for experiment.
3. 424 kinds of component maps are made for these chosen genes.

Finally, by the SOM method, the genes are clustered according to the following procedures.

- (a) The data vector is made based on dark and bright informations on each unit on the component map. In the case, we use the vector of 100 dimensions because the map size was  $10 \times 10$ .
- (b) ID of the chosen gene is assumed to be a teacher label of each item.
  - The number of data is 424, and each data has a multi-dimensional vector of 100 dimensions.
- (c) The result by SOM learning is shown in the Fig. 2.

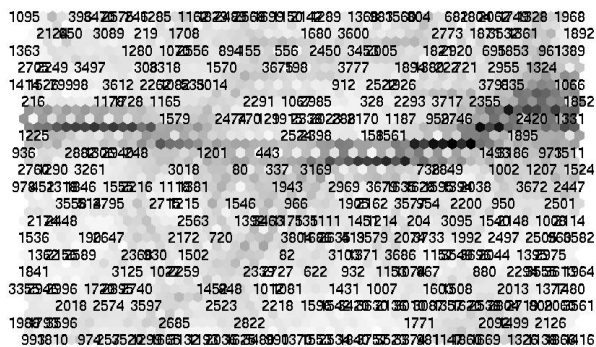


Figure 2: Clustering of genes.

Here, a plane SOM and a spherical SOM were used for the DNA microarray analysis. And, the result was equivalent with a usual clustering method. However, the result was obtained in a reasonable shorter time.

## References

- [1] Knudsen, S., *A Biologist's Guide to Analysis of DNA Microarray Data*, Wiley-Liss, 2002.
- [2] Kohonen, T., *Self-Organizing Maps*, Springer, 1995.