Suitability of Spherical SOM for Gene Expression Analysis

Hirokazu Nishio, [1] **Ken-nosuke Wada,** [2] **Yoshiko Wada,**[2]
Md. Altaf-Ul-Amin,[1] **Shigehiko Kanaya**[1]

1   Introduction.

Self-Organizing Maps(SOM)[1] are visualization methods by creating dimension-reduced feature maps, where units in low-dimensional represantaion space are associated with those in original high-dimentional data space. In the present study, we consider two-types of SOMs, one has planar representation space called 'Plain SOM' and the other has spherical space called 'Spherical SOM'[2]. In general, the profile vectors of gene expression are normalized to unity in length in order to focus not on the absolute quantity of expression but the similarity of the direction[3]. Since the normalized vectors are distributed on the surface of a supersphere, it tempts to think that the Spherical SOM is suitable for the data of gene expression. But this idea had no ground because the surface of a sphere is topologically different from that of a supersphere. In the present study, we propose the measure for suitablity of SOM to a given data set, and confirmed that the Spherical SOM was actually suitable for the data of gene expression.

2   Method.

If a SOM has the representation space which can express the structure of data without distorting, distance between units in representation space should be proportional to that in original data space. This is quantified by the linearity of 'd-d plot' as follows: All pairs of the units are plotted in the two-dimensional plane. Horizontal axis corresponds to the distance in representation space, and the vertical axis corresponds to the distance in data space. The linearity of a distribution is quantified by the coefficient of determination of regression analysis, which is called 'N-measure'.

3   Result and Discussion.

In order to demonstrate the validity of N-measure, we examine three data sets as follows:

- Randomly distributed 1800 points on a plain surface consisting of two axes. Variance of each axis is unity.

- Randomly distributed points on a sphere surface consisting of three axes. Initially three dimensional 1800 vectors are randomly generated. Here, variance of each axis is unity. Then all the vectors are normalized to unity in length.

- Actual 8-dimensional normalized gene expression profiles of *B. subtilis.*

---

[1]Nara Institute of Science and Technology, Graduate School of Information Science, 8916-5 Takayama, Ikoma, Nara 630-0101. E-mail: `hiroka-n@is.aist-nara.ac.jp`
[2]UNTROD Inc. 4-4-123 Aoyama, Narashi, Nara-ken, 630-8101

Figure 1 shows the d-d plots for those data with Plain SOM (upper) and Spherical SOM (lower). In case of plain surface data, a linear relationship is observed in the d-d plot of Plain SOM (left upper). The N-measure of Plain SOM (0.8901: see Table 1) is greater than that of Spherical SOM (0.3212). On the other hand, in case of sphere surface data, a linear relationship is observed in the d-d plot of Spherical SOM (middle lower). The N-measure of Plain SOM (0.3782) is less than that of Spherical SOM (0.9764). Thus, suitablity of SOM to a given data set can be estimated by N-measure. The N-measure increases when a SOM has the representation space which can express the structure of data without distorting. In case of actual 8-dimensional gene expression data, the N-measure of Plain SOM (0.1682) is less than that of Spherical SOM (0.6044). We conclude the Spherical SOM is suitable for the analysis of the normalized data of gene expression.
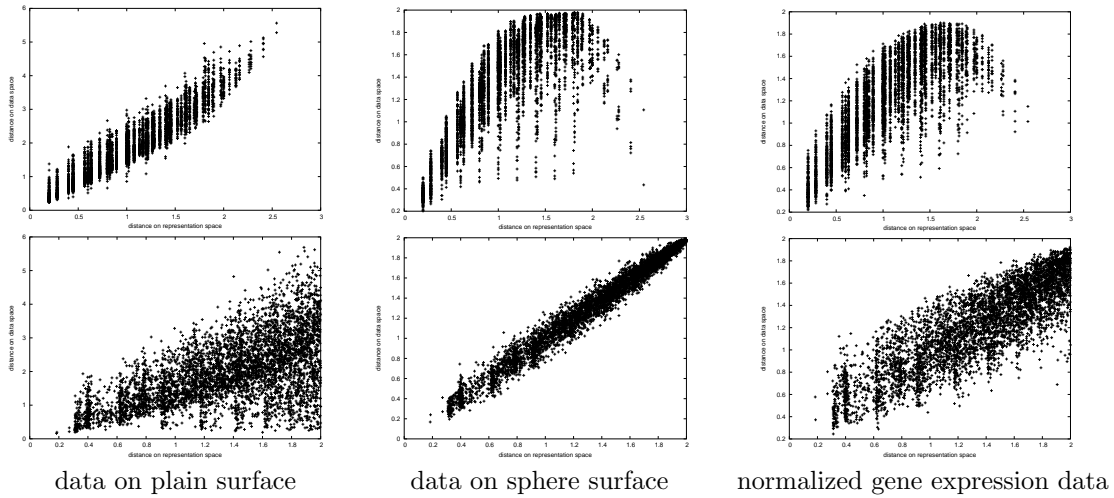

| data on plain surface | data on sphere surface | normalized gene expression data |

Figure 1: d-d plot(upper:Plain SOM, lower:Spherical SOM)

	Plain SOM	Spherical SOM
data on plain surface	0.8901	0.3212
data on sphere surface	0.3782	0.9764
data of normalized gene expression	0.1682	0.6044

Table 1: coefficient of determination

References

[1] Teuvo Kohonen. Self-Organizing Maps. Springer Series in Information Sciences. Springer, second edition, 1997.

[2] Helge Ritter. Self-Organizing Maps on non-euclidean Spaces. Kohonen Maps, pp. 97-108, ed. E. Oja and S. Kaski, 1999

[3] Hirokazu Nishio, Md. Altaf-Ul-Amin, Tetsuo Sato, Ken-nosuke Wada, Yoshiko Wada, Kotaro Minato, Kazuo Kobayashi, Naotake Ogasawara, Shigehiko Kanaya. Visualization of Gene Classification Based on Expression Profile Using BL-SOM. Proc. WSOM'03, pp. 101-106, 2003.