2048
doi:10.1017/S1431927618010723
Microsc. Microanal. 24 (Suppl 1), 2018
© Microscopy Society of America 2018

Development and Application of an Automatic Cluster Analysis System for EPMA Using Neural Network

Norihisa Mori, Naoki Kato and Masaki Morita.

[1.] JEOL Ltd., 1-2 Musashino, 3 Akishima, Tokyo 196-8558 JAPAN

Machine learning technology using a neural network such as deep learning has made progress in recent years. It's the feature that the neural network is able to get a solution near human's sense. It's possible to apply this function to search of phases in EPMA and EDS mapping data. It's able to use a statistical analysis and the technique of PCA in the past in case of the small data or the simple component. But for an analysis to become complicated when handling microstructures, trace elements, a lot of dimensions and big amount of data, this method weren't enough. Therefore a human went to a judgement of a presence of the phases mainly. It's possible to judge this human's sense for machine learning by neural networks. When using neural networks, it's a problem how to give reproducibility and a threshold.
For these problems, the high-speed cluster analyses (HSCA) using the hierarchy cluster analyses (HCA) [1]which is a classic method and self-organization method (SOM) [2] which is mechanical learning by neural networks have been developed in this research. (Figure 1) HSCA is the automatic phase analysis by unsupervised learning for EPMA mapping data.

Our purpose is to express map data of high dimensional (multi elements) EPMA by one phase map using multivariate analysis. We have tried the automatic cluster analysis by using the principal component analysis (PCA). [3] But PCA does not suit an analysis of a sample which includes trace elements and minor elements. For such a sample, HCA would be obtained a good result while there is a large amount of calculation. Data of EPMA has a capacity with 10,000,000 pixels from 10,000 pixels. It's the too big capacity to process it in HCA. We reduced computing time by using hierarchical SOM.
SOM used for only cluster-ization of first step because reproducibility of data was low.
After doing enough data small by SOM, a hierarchy cluster analyses is performed by a second-stage.
The bad influence which are a problem in time of HCA and reproducibility of SOM by 2 steps of this processing, reduction was achieved.

As for HSCA, First calculation is to compare partially. It is divided into about 30 to 1000 sub-clusters by SOM. This sub-cluster is quantized by the size of the statistical error by SOM. These sub-clusters are divided into the phases of about 4 to 64 by HCA. This HCA is using ward method. [4] HSCA decides the approximate number of the phase by the threshold value using statistical error of the scalar amount of all elements. HSCA can calculate this without setting up any default parameters by operator. And this can output many cluster information such as X-ray intensity, mass percent, area, area ratio and C/ V value of each clusters. It is possible to edit the result of HSCA using a conventional scatter diagram analysis.

Phase map using HSCA was made in the Zn coated steel sheet which did etching test. This steel plate made them corrode in salt water for 1000 hours. This corrosion sample includes 8 kinds of element. Experience is needed to understand 8 of map data. Figure 2b is the result by which a cluster analyzed these data by PCA and manual operation. Figure 2a is automatically cluster analyses by HSCA. It was possible to indicate the corrosion look in a phase map by using HSCA without depending on special knowledge and experience.

https://doi.org/10.1017/S1431927618010723 Published online by Cambridge University Press

References:

[1] Lance, G. N and Williams 1967 Computer Journal **9**, 380-383
[2] Teuvo Kohonen "*Selfe-Organizing maps*" (Springer-Verlag, Espoo, FINLAND)
[3] Norihisa M. 2015 in: *Book of Abstracts of the European Microbeam Analysis Society 14th European Workshop.* 374
[4] Ward, J.H.Jr. 1963 Journal of the Americam Statistical Association, **58**, 236-244

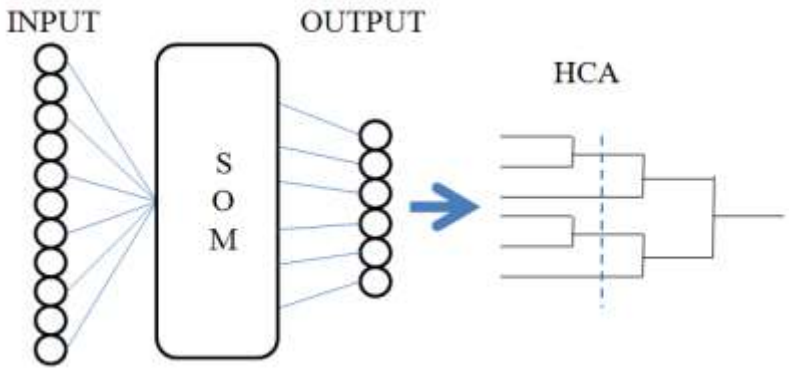

Figure 1. HSCA outputs a sub-cluster by SOM first. Next this sub-cluster is analyzed by HCA.

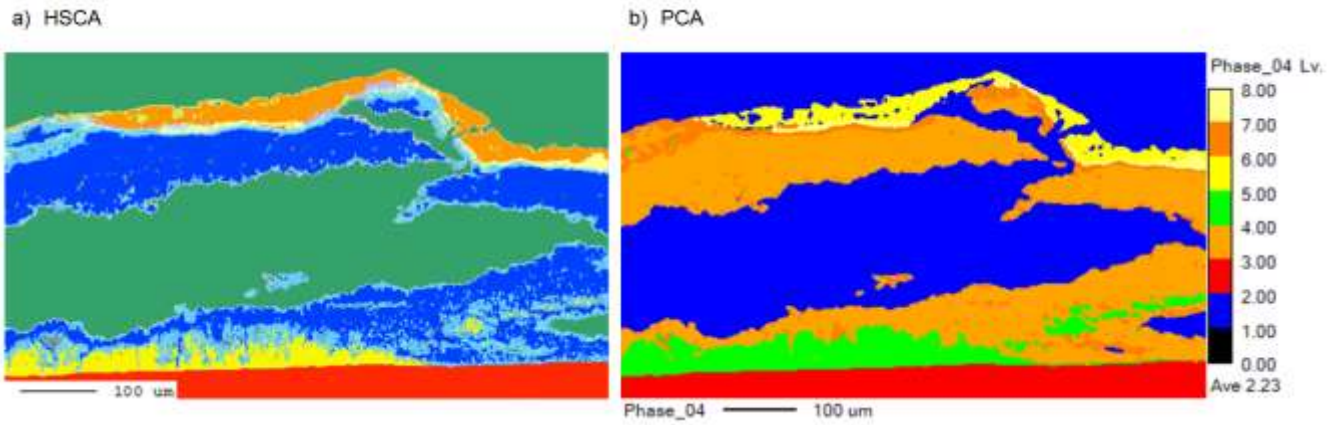

Figure 2. Comparison of a) the result of the unsupervised learning of HSCA and b) the result of manual analyzed using PCA.

https://doi.org/10.1017/S1431927618010723 Published online by Cambridge University Press