BIO Web of Conferences **129**, 10043 (2024) https://doi.org/10.1051/bioconf/202412910043
EMC 2024

Interpretable evaluation of STEM images of nanostructures via homology analysis

Mr. Ryuto Eguchi[1,2], Dr. Yu Wen[1,2], Dr. Ayako Hashimoto[1,2]
[1]National Institute for Materials Science, Tsukuba, Japan, [2]University of Tsukuba, Tsukuba, Japan

Background incl. aims

Recently, a mathematical framework called persistent homology (PH) has made it possible to quantify materials structural information at a wide range of scales, from the atomic to the nanoscale. PH quantitatively expresses the hole structures of data in terms of their number and scale. Particularly, it is highly suited for discerning the subtle order within inhomogeneous structures, such as amorphous [1] and glass materials [2]. Until now, PH analysis has predominantly focused on three-dimensional atomic arrangements; however, it is also feasible on two-dimensional image data. Here, we applied PH analysis for two-dimensional TEM images, representing a highly useful approach in structural analysis.

In our previous works [3,4], the homological feature known as 'Betti number' was applied for structures of self-assembled Pt/CeO_2 nanocomposites, which were captured by scanning TEM (STEM). The N-th Betti number corresponds to the number of N-dimensional holes, such as connected components, rings (0- and 1-dimensional holes, respectively), and so forth. This homological feature could successfully quantify CeO_2 phase connectivity and further, relationship with the oxygen ion conductivity.

To explore more effective descriptor for the nanostructures, we apply one of the most used PH methods, Persistent Diagram (PD). The key concept in PD lies in tracking the scale required for the appearance (birth, b) and disappearance (death, d) of the N-dimensional holes by continuous deformation of object which called ‘filtration’. Consequently, it includes information on the shapes of the N-dimensional holes, unlike the Betti number. We aim to demonstrate its effectiveness for nano-structural analysis and extract important homological feature for classifying the Pt/CeO_2 nanostructures. To ensure both quantitativity and interpretability, we employed a consistent approach that directly extract interpretable features from PDs.

Method

Firstly, Pt/CeO_2 nanocomposites were synthesized by the annealing of the Pt_5Ce alloy. The 12 nanocomposites with various nanostructures were prepared by changing annealing temperature (500, 600, and 700°C) and syngas ratio (CO:O2 = 0:1, 1:1, 2:1, and 3:1). The nanostructures were characterized through STEM (JEM-2100F, JEOL, Japan) operating at an acceleration voltage of 200 kV. Then, the obtained images were binarized and noise-removed with the OpenCV library in Python. The sequential procedures related to PD acquisition, vectorization for Principal Component Analysis (PCA) were conducted by using data analysis software “Homcloud” [5]. Finally, we used random forest method to find the most important descriptor for classifying Pt/CeO_2 nanostructures.

Result

The binarized STEM images clearly show the self-assembled Pt/CeO_2 nanocomposites that consist of Pt (white) and CeO_2 (black) phases. We focused on the CeO_2 phase for PH analysis to examine its relationship with oxygen ion conductivity. The nanostructures changed from a maze-like to a striped appearance as the annealing temperature increased. Their 0th and 1st PDs also changed depending on the structural changes.

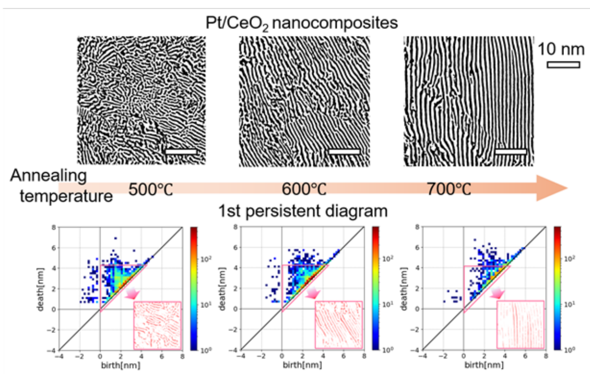
To clarify the relationship between the structural changes and distribution changes of b-d points in PDs, we extracted five interpretable features, each based on the understanding of individual quadrants in the 0th and 1st PD. The three features, the average width and total length of the striped CeO_2 phases, and the number of CeO_2 phases can be obtained from the 0th PDs. Their trends toward annealing temperature coincide with those from the actual STEM images. The number of ring and gulf-like structures can be obtained from the 1st PDs, focusing on negative and positive b region, respectively. These quantification with PDs could capture trends more clearly compared with conventional observation.

Furthermore, we conducted PCA with vectorized PDs to extract their critical information, which emphasizes the difference between homological features in structures. By the first and second principal components in the 0th and 1st PDs, the 12 nanostructures were relatively well categorized. Through PD reconstruction using the first principal components in the 0th and 1st PDs, we identified a critical region in the PDs: the region in the 1st PDs with d value smaller than characteristic size. In this critical region, d value corresponds to the size of the CeO_2 gulf-like phases. This result suggests that the number of small gulf-like phases is effective as a simple interpretable feature to differentiate Pt/CeO_2 nanostructures. Finally, we applied all interpretable features extracted from PDs thus far to a random forest classification and evaluated their importance. As a result, two key descriptors emerged: the width of the CeO_2 phase and the number of small gulfs. Remarkably, in the scatter plot of the two descriptors, the 12 nanostructures could be classified effectively. In this manner, using a few simple interpretable descriptors, we can more easily discuss and quantitatively evaluate the structural differences arising from variation in synthesis conditions, compared to PCA.

Conclusion

This study investigated the effectiveness of PH analysis in analyzing the STEM images of the nanostructures. Firstly, five interpretable features could be extracted directly from the 0th (the average width and total length of the striped CeO_2 phases, and the number of CeO_2 phases) and 1st (the number of ring and gulf-like structures of CeO_2 phase) PDs. Regarding the gulf-like structure, the PCA results suggest that the number of smaller structures than the characteristic size could particularly differentiate the 12 nanostructures. Finally, we showed that key descriptors can classify nanostructures through the random forest classification, enabling us to interpret their differences easily.

Graphic:



Binarized HAADF-STEM images of Pt/CeO_2 nanostructures synthesized at different temperature (top), and 1st PDs of their CeO_2 phases (bottom). The triangle regions in the 1st PDs indicate critical regions for classifying the nanostructures. Images insets show the result of inverse analysis on each Pt/CeO_2 nanostructures obtained from their triangle regions in the1st PDs.

Keywords:

persistent homology, interpretable machine learning

Reference:

[1] Y. Hiraoka, T. Nakamura, A. Hirata, E.G. Escolor, K. Matsue and Y. Nishiura, PNAS 113, 7035 (2016).
[2] T. Nakamura, Y. Hiraoka, A. Hirata, E.G. Escolar and Y. Nishiura, Nanotechnol. 26, 304001 (2015).
[3] Y. Wen, A. Hashimoto, A.S.B.M. Najib, A. Hirata and H. Abe, Appl. Phys. Lett. 118, 054102 (2021).
[4] Y. Wen, H. Abe, A. Hirata and A. Hashimoto, ACS Appl. Nano Mater. 4, 13602 (2021).
[5] Homcloud : https://homcloud.dev/