Using a Self Organising Map for Clustering of Atomistic Samples

Aquistapace F.

Facultad de Ciencias Exactas y Naturales, UNCuyo

September 10, 2021

Introduction

This software implements a Self Organising Map (also known as a Kohonen Network) for clustering of atomistic samples through unsupervised learning. It is written in Python 3.8.8 and depends on the external packages NumPy (version 1.20.1) and Pandas (version 1.2.4). The core of the algorithm is a neural network composed of an input layer and an output layer, where the number of output neurons determines the number of groups the atoms are going to be classified into.

V. 1

Description:

The first version of this software implements a command-line interface through which the user can select the input file, choose which features of the data to use and set the parameters of the algorithm. This interface also let's the user know the actual status of the program, and the time elapsed when it is finished.

An immediate problem of this version is the normalization method in the pre-processing stage of the analysis. For every column C, associated with a selected feature, the normalization is given by Eq. 1:

$$C_{norm} = \frac{C}{C_{max}} \tag{1}$$

Where C_{norm} is the normalized data column and C_{max} is the maximum value of the original column. This is not the ideal normalization method, since the range of the new column is now $[C_{min}/C_{max}, 1]$ (where C_{min} is the minimum value of C), instead of being $[0, 1]$. This issue is going to be fixed in the next version of the software.

Testing:

BCC Fe Bulk With a Void:

As a simple test, a 25×10^4 atoms BCC Fe bulk with a void was analysed. The sample was created and relaxed with LAMMPS. Since the sample has no defects, the goal was to use the SOM to identify the void in the center of the simulation box. This was achieved using parameters $\sigma = 1$, $\eta = 0.5$ and $f = 1$, clustering the atoms into 3 groups and using the centro-symmetry parameter and the coordination (both calculated with LAMMPS) as features.

The results are shown in Fig. 1 and Fig. 3. Although the algorithm classified the atoms into 3 groups, no atom was assigned to group 1. For the previously specified parameters, the SOM didn't produce useful results when clustering the atoms into only 2 groups. For V.1 of the algorithm, this analysis took around $26.4s$ to complete.

HEA Nano-foam Under Compression:

Fig. 5 shows the result of applying the SOM to a compressed HEA nano-foam sample, with the goal of classifying the atoms into 2 groups. The network was trained using parameters $\sigma = 1$, $\eta = 0.5$ and $f = 1$. The features used in this occasion were:

- Centro-symmetry parameter
- Coordination, via Coordination Analysis with cutoff radius $r_c = 6$
- Atomic volume, via Voronoi Analysis

On the other hand, Fig. 6 shows the same sample, with the atoms classified by structure type (FCC, HCP, BCC and Other) using the PTM algorithm. It can be observed that the yellow atoms in Fig. 5 correspond to the HCP, BCC and Other atoms in Fig. 6. While the atoms in the remaining group of the SOM classification correspond to the FCC atoms of the sample. In this sense, the SOM is correctly identifying the atoms associated with defects and/or that belong to the surface of the sample.

A problem that has been noticed while performing this test is that the order in which the features are specified seems to affect the resulting classification of the atoms. In this case, the algorithm produced the results shown in Fig 5 when the features were specified in the order *coordination* $\rightarrow$ *centro-symmetry* $\rightarrow$ *atomic volume*, but couldn't get the same results when the order of the features was *coordination* $\rightarrow$ *atomic volume* $\rightarrow$ *centro-symmetry*. This could be solved in future updates of the algorithm.

A summary of all the tests performed is presented in Table 1.

Test	**Nº of atoms**	f	σ	η	**Groups used**	**Time elapsed**
BCC Fe with void	$\approx 25 \times 10^4$	1	1	0.5	3	$26.4s$
HEA Nanofoam under compression	$\approx 20 \times 10^5$	1	1	0.5	2	$197.7s$

Table 1: Size of the sample, parameters and performance of V.1 of the algorithm for every test.

V. 2

Fixes

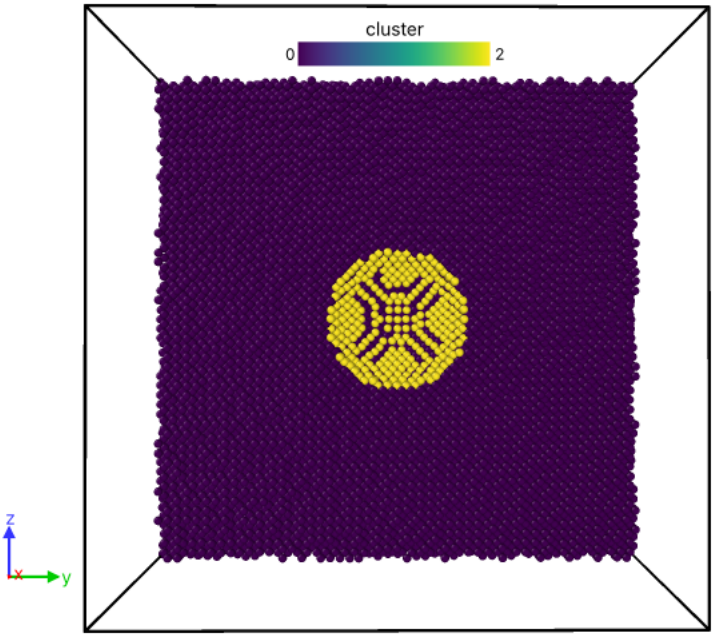

Figure 1: BCC Fe bulk with void. The atoms have been clustered into 3 groups using parameters $\sigma = 1$, $\eta = 0.5$ and $f = 1$. The centro-symmetry and coordination (via Coordination Analysis with $r_c = 6$) were used as features. A slice of the sample, with groups 0 (purple) and 2 (yellow), is shown.

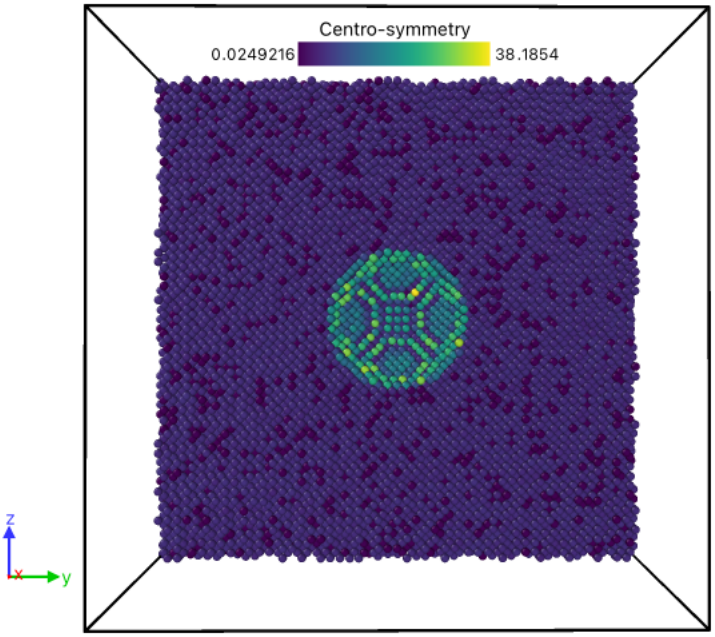

Figure 2: BCC Fe bulk with void. The atoms have been clustered using the centro-symmetry parameter (as calculated by LAMMPS). A slice of the sample is shown.

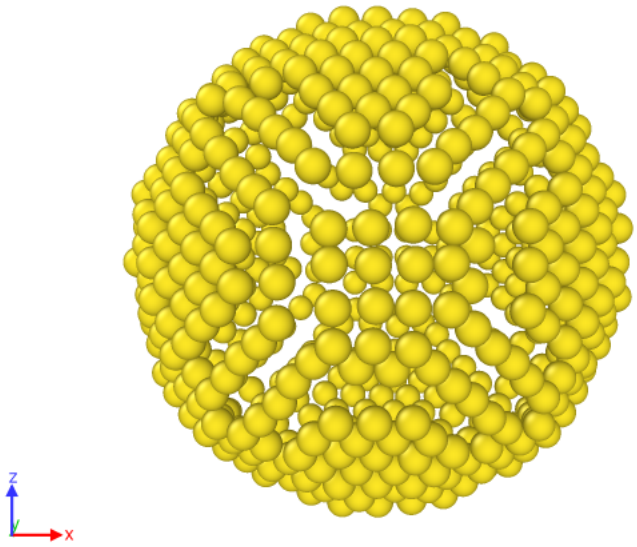

Figure 3: Void of the BCC Fe bulk sample. The atoms belonging to group 2, as classified by the SOM, are shown. The centro-symmetry and coordination (via Coordination Analysis with $r_c = 6$) were used as features.

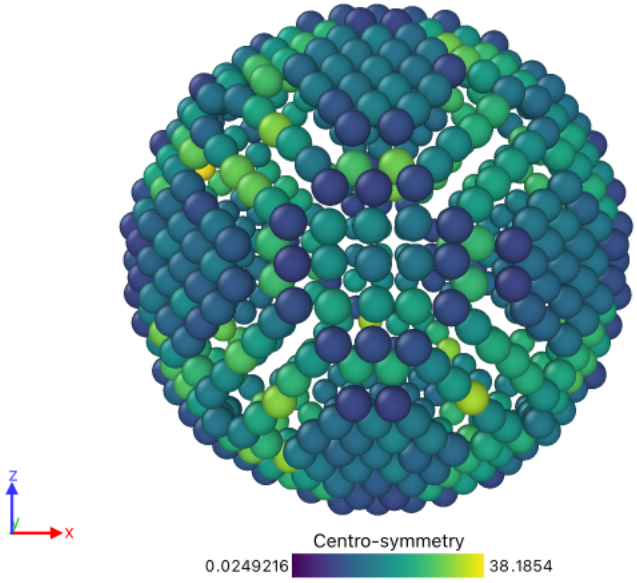

Figure 4: Void of the BCC Fe bulk sample. The atoms have been clustered using the centro-symmetry parameter (as calculated by LAMMPS).

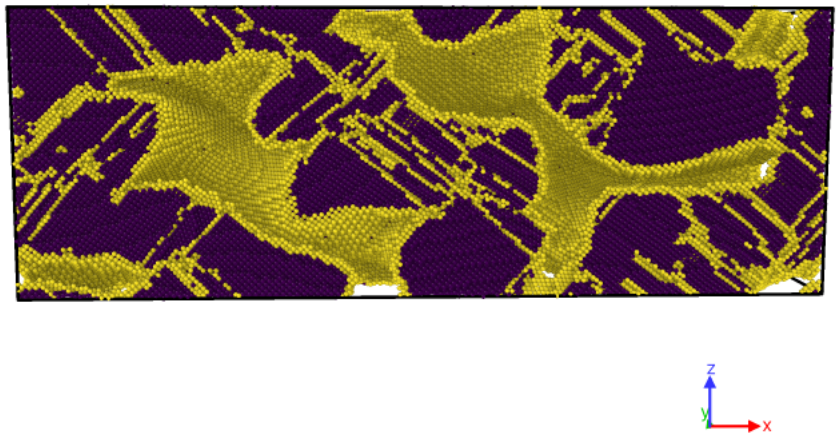

Figure 5: HEA Nano-foam under compression. The atoms have been clustered into 2 groups using parameters $\sigma = 1$, $\eta = 0.5$ and $f = 1$. The centro-symmetry, coordination (via Coordination Analysis with $r_c = 6$) and atomic volume (via Voronoi Analysis) were used as features.

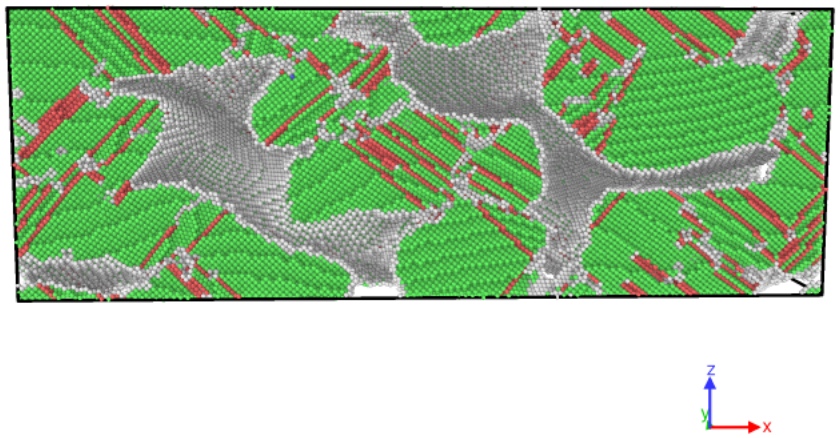

Figure 6: HEA Nano-foam under compression. The atoms have been clustered using the PTM algorithm into 4 categories: FCC (green), HCP (red), BCC (purple) and Other (white).