Fuzzy c-Means Clustering for Persistence Diagrams

Thomas Davies[1], Jack Aspinall[2], Bryan Wilder[3], and Long Tran-Thanh[4]

[1]University of Southampton, [2]University of Oxford, [3]Harvard University, [4]University of Warwick

Motivation

- Most current approaches to integrating topological information into machine learning implicitly map persistence diagrams to a Hilbert space, resulting in deformation of the underlying metric structure whilst also generally requiring prior knowledge about the true topology of the space.
- In this paper we give an algorithm for Fuzzy c-Means (FCM) clustering directly on the space of persistence diagrams, enabling unsupervised learning that automatically captures the topological structure of data, with no prior knowledge or additional processing of persistence diagrams.
- We prove the same convergence guarantees as traditional FCM clustering: every convergent subsequence of iterates tends to a local minimum or saddle point.
- We run experiments where the fuzzy nature of our topological clustering is capitalised on: lattice structure classification in materials science and pre-trained model selection in machine learning.

Fuzzy clustering persistence diagrams

We extend the FCM algorithm originally proposed by [1]. Our r_{jk} is the same as traditional FCM clustering, adapted with the Wasserstein distance. That is,

$$r_{jk} = \left(\sum_{l=1}^{c} \frac{W_2(\mathbb{M}_k, \mathbb{D}_j)}{W_2(\mathbb{M}_l, \mathbb{D}_j)} \right)^{-1}. \tag{1}$$

To update $\mathbb{M}_k$, we compute the weighted Fréchet mean $\hat{\mathbb{D}}$ of the persistence diagrams $\{\mathbb{D}_j\}_{j=1}^n$ with the weights $\{r_{jk}^2\}_{j=1}^n$. Specifically,

$$\mathbb{M}_k \longleftarrow \arg\min_{\mathbb{D}} \sum_{j=1}^{n} r_{jk}^2 W_2(\hat{\mathbb{D}}, \mathbb{D}_j)^2, \text{ for } k = 1, \ldots, c. \tag{2}$$

As the weighted Fréchet mean extends weighted centroids to general metric spaces, this gives our fuzzy cluster centres. By alternatively updating (1) and (2) we get a sequence of iterates. Theorem 1 provides the same convergence guarantees as traditional FCM clustering.

Theorem 1. Every convergent subsequence of the sequence of iterates obtained by alternatively updating membership values and cluster centres with (1) and (2) tends to a local minimum or saddle point of the cost function $J(R, \mathbb{M}) = \sum_{j=1}^{n} \sum_{k=1}^{c} r_{jk}^2 W_2(\mathbb{M}_k, \mathbb{D}_j)^2$.

Computing the weighted Fréchet mean

Turner et al. [2] give an algorithm for the computation of Fréchet means. We extend their algorithm and proof of convergence to the weighted case.

We start by finding matched points

$$\left[x_j^{(i)}\right]_{i=1}^{m} \longleftarrow \text{Hungarian}(\mathbb{M}_k, \mathbb{D}_j), \text{ for each } j = 1, \ldots, n. \tag{3}$$

Partition $1, \ldots, n$ into the indices of the off-diagonal points $\mathscr{J}_{\text{OD}}^{(i)}$ and the indices of the diagonal points $\mathscr{J}_{\text{D}}^{(i)}$. If $\mathscr{J}_{\text{OD}}^{(i)} = \emptyset$, then $y^{(i)}$ is a copy of the diagonal. Otherwise, let w be the weighted mean of the off-diagonal points and w_Δ be the point on the diagonal closest to w. Then our update is

$$y^{(i)} \longleftarrow \frac{\sum_{j \in \mathscr{J}_{\text{OD}}^{(i)}} r_{jk}^2 x_j^{(i)} + \sum_{j \in \mathscr{J}_{\text{D}}^{(i)}} r_{jk}^2 w_\Delta}{\sum_{j=1}^{n} r_{jk}^2}, \text{ for } i = 1, \ldots, m. \tag{4}$$

We alternate between (3) and (4) until the matching remains the same. Theorem 2 proves that this algorithm converges to a local minimum of the Fréchet function.

Theorem 2. Given diagrams $\mathbb{D}_j$, membership values r_{jk}, and the Fréchet function $F(\hat{\mathbb{D}}) = \sum_{j=1}^{n} r_{jk}^2 W_2(\hat{\mathbb{D}}, \mathbb{D}_j)^2$, then $\mathbb{M}_k = \{y^{(i)}\}_{i=1}^{m}$ is a local minimum of F if and only if there is a unique optimal pairing from $\mathbb{M}_k$ to each of the $\mathbb{D}_j$ and each $y^{(i)}$ is updated via (4).

Experiments: Synthetic data

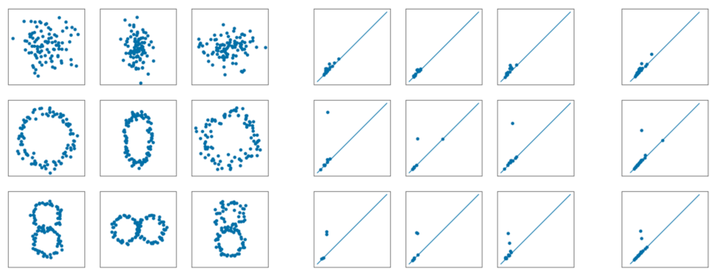

Figure 1: Given datasets with zero, one, or two holes (left), we cluster their persistence diagrams (middle), resulting in cluster centres (right) with zero, one, or two off-diagonal points: our cluster centres reflect the topology of the datasets.

Experiments: Lattice structures

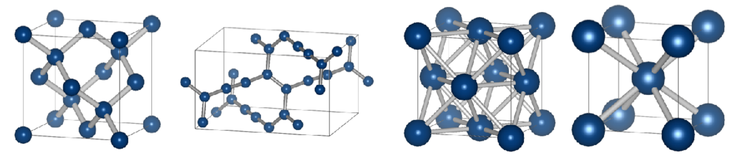

Figure 2: Our algorithm is able to cluster transformed carbon allotropes (left two) and FCC/BCC structures (right two) from materials science where comparable Wasserstein barycentre clustering algorithms fail.

A key property for machine learning in materials science has been identified as "invariance to the basis symmetries of physics [...] rotation, reflection, translation" [3]. The properties of persistence diagrams mean that we can successfully cluster the atomic coordinates derived from the same base unit-cell regardless of the transformations applied to the coordinate system, fulfilling the key property identified above.

Experiments: Decision boundaries

We cluster the 1-persistence diagram of the decision boundaries of models and tasks. Our clustering successfully captures information about the performance of a model on potentially unseen tasks.

Table 1: Performance increase/decrease over average task performance when using learnt fuzzy membership values for model selection. The increase in performance demonstrates that our fuzzy clustering automatically clusters models near tasks they perform well on.

	Performance change vs random model selection (%)		
	Top-3	Top-2	Top-1
MNIST	$+6.17^{\pm 2.18}$	$+10.81^{\pm 1.88}$	$+20.88^{\pm 4.08}$
FashionMNIST	$+16.46^{\pm 4.00}$	$+21.94^{\pm 4.73}$	$+23.30^{\pm 8.72}$
Kuzushiji	$+6.61^{\pm 1.78}$	$+11.18^{\pm 2.45}$	$+21.89^{\pm 5.54}$

References

[1] J. C. Bezdek, "A convergence theorem for the fuzzy isodata clustering algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, no. 1, pp. 1–8, Jan. 1980, ISSN: 1939-3539. DOI: 10.1109/TPAMI.1980.4766964.

[2] K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer, "Fréchet means for distributions of persistence diagrams," *Discrete & Computational Geometry*, vol. 52, pp. 44–70, 2012.

[3] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent advances and applications of machine learning in solid-state materials science," *npj Computational Materials*, vol. 5, no. 1, p. 83, 2019, ISSN: 2057-3960. DOI: 10.1038/s41524-019-0221-0. [Online]. Available: https://doi.org/10.1038/s41524-019-0221-0.