Model Transferability Informed by Embedding’s Topology

Felipe Gutiérrez figutier@uc.cl
Hans Lobel halobel@uc.cl

Pontificia Universidad Católica de Chile

Motivation

The Problem: The rapid proliferation of large-scale model zoos makes it difficult to select the optimal pre-trained backbone for downstream tasks.

The Bottleneck: Exhaustively fine-tuning every available model to find the best one is computationally prohibitive.

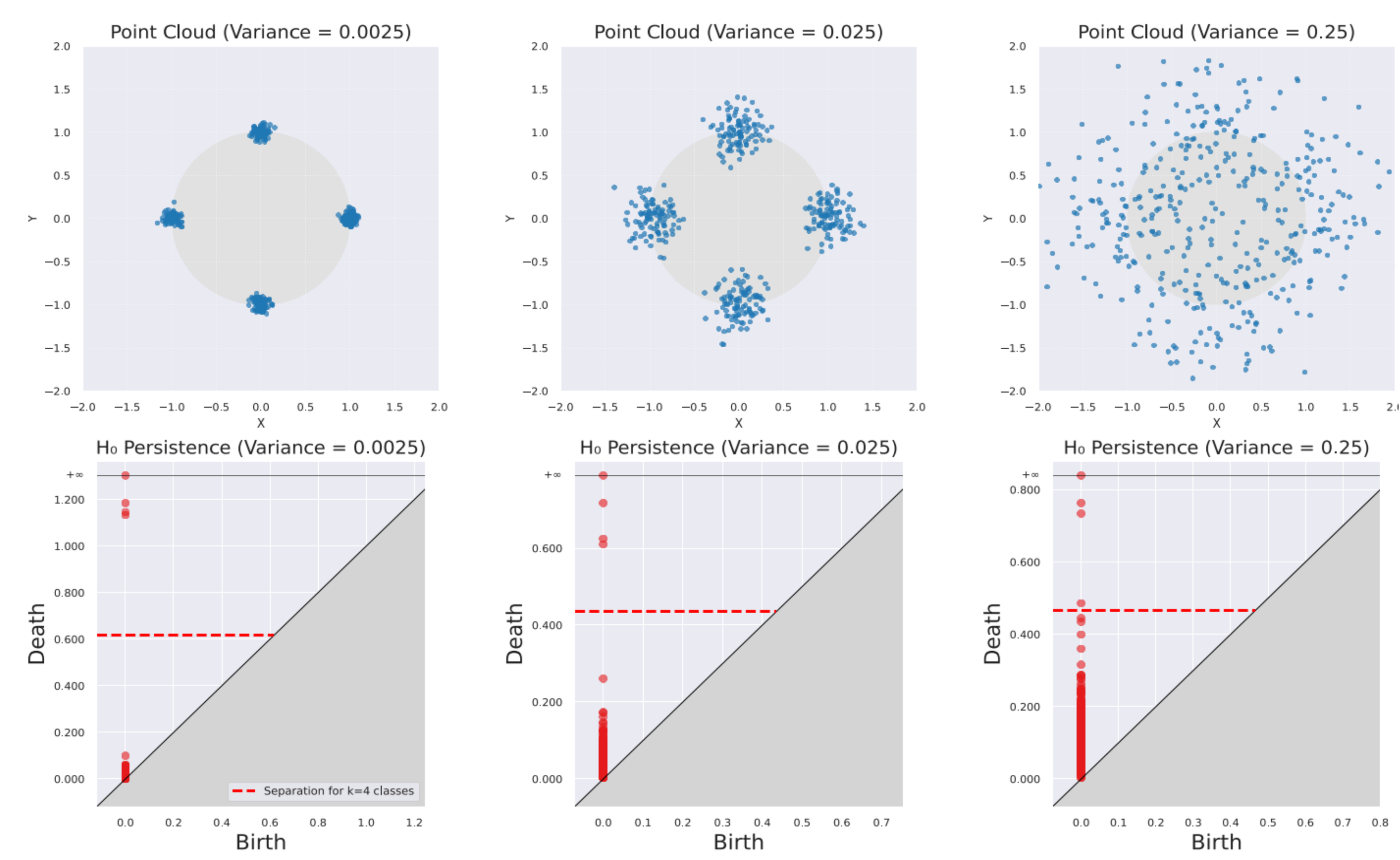
The Goal: Heuristic to predict which models will perform better after transfer *without* performing the expensive fine-tuning process.

Link to paper

Geometric Insights

Neural Collapse: The cluster structure of a neural network's embeddings generated at late training on a classification problem. Members of the different classes tend to separate from each other, as members of the same class tend to converge to the mean of the class.

Topological Signature: Reveals this structure on the 0-dimensional homology groups of the persistence diagrams of the embeddings. As seen in the image, the number of clusters is revealed as high persistence homology groups, while the rest have persistence values close to 0.



Persistent diagrams of cloud points of 4 clusters sampled from a circle at different variances.

Our paper

Given the expected diagram structure on mature embeddings, we define:

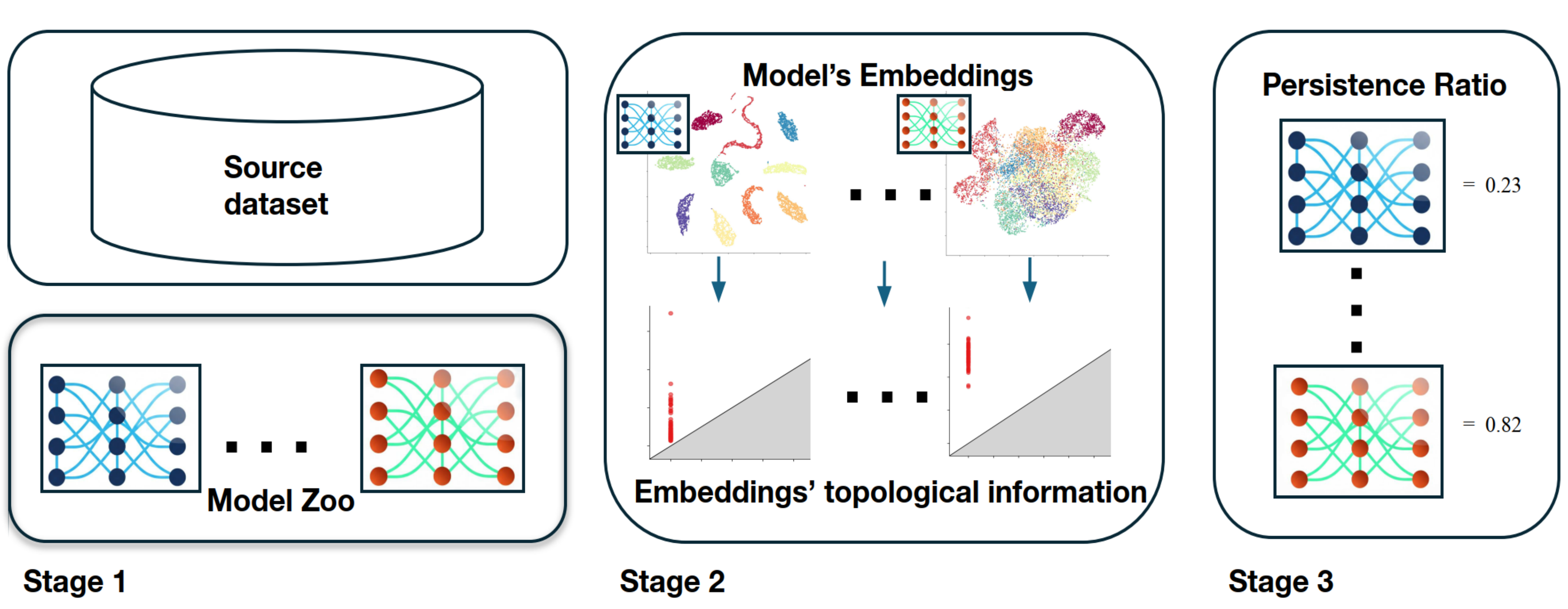
Inter-class Separation: $H_{sep} = \frac{1}{k-1}\sum_{i=1}^{k-1} p_i$

Intra-class Tightness: $H_{tight} = \frac{1}{n-(k-1)}\sum_{i=k}^{n} p_i$

Persistence Ratio (PR): $H_{ratio} = \frac{H_{tight}}{H_{sep}}$

Where n is the number of homology modules, k is the number of classes and p_i is the persistence value of the i_{th} 0-dimensional homology module ordered by decreasing persistence value.

Methodology



Stage 1: Given a model zoo and a dataset, we extract the embeddings learned by each pretrained model.

Stage 2: The Vietoris-Rips filtration of the embeddings is computed, and the persistent diagram for each model is generated.

Stage 3: H_ratio is calculated, and the models are ranked according to the additive inverse of the value.

Metrics: We use Kendall’s Tau to verify the ranking correlation of $-H\_ratio$ with the performance of the models after fine tuning

Results

Performance: Negative PR of the original embeddings excels at predicting high-performing models on fine-tuning tasks.

Reliability: Overall, we find that our measure, used on the original model, is not only a high-performing heuristic, but a very reliable one compared to other measures, due to its consistency across datasets.

Source vs. Target embeddings: While the topological cluster structure on the target dataset performs well on some datasets, it is highly unreliable due to low or even inverse relations on other datasets.

Dataset	LEEP	NCE	LogME	NC Score	PR_target	PR
Aircraft	0.289	0.511	0.36	**0.644**	**0.644**	0.556
Birdsnap	0.333	0.644	0.467	**0.778**	0.733	0.689
Caltech 101	0.422	**0.733**	0.584	0.422	**0.733**	0.689
StanfordCars	0.424	0.566	0.519	**0.801**	0.283	0.66
CIFAR 10	0.477	0.159	**0.705**	0.205	0.341	0.614
CIFAR 100	0.422	0.111	0.629	0.244	-0.067	**0.644**
DTD	0.022	-0.289	0.494	0.2	-0.556	**0.689**
Pets	0.467	**0.733**	0.511	0.556	0.378	0.556
SUN397	0.449	0.764	0.719	0.315	0.405	**0.809**
MEAN	0.367	0.437	0.554	0.463	0.322	**0.656**

Kendal Tau results for model zoo on each dataset tested

Conclusions

Topology describes late-stage training structure: We show that persistent homology effectively quantifies the geometric evolution of embeddings, capturing the dynamics of class separation and cohesion.

Transferability metric: Our proposed measure outperforms established baselines in ranking pre-trained models on classification tasks, proving to be a robust and reliable metric.

Intrinsic structure is predictive: Our experiments show that the topological structure of the source representation is a stronger predictor of transferability than target-based heuristics.

Acknowledgements

The authors acknowledge the support of the National Agency for Research and Development ANID-Subdirección de Capital Humano/Doctorado Nacional/2024-21241334, the National Center of Artificial Intelligence (ANID grant FB210017), the Millennium Institute for Foundational Research on Data (ANID grant ICN17_002), and FONDECYT grant 1241772.