Wasserstein Auto-Encoders of Merge Trees (and Persistence Diagrams)

Mathieu Pont*

CNRS, LIP6 – Sorbonne Universités, UPMC, CNRS – France

Résumé

This paper presents a computational framework for the Wasserstein auto-encoding of merge trees (MT-WAE), a novel extension of the classical auto-encoder neural network architecture to the Wasserstein metric space of merge trees. In contrast to traditional auto-encoders which operate on vectorized data, our formulation explicitly manipulates merge trees on their associated metric space at each layer of the network, resulting in superior accuracy and interpretability. Our novel neural network approach can be interpreted as a non-linear generalization of previous linear attempts (9) at merge tree encoding. It also trivially extends to persistence diagrams. Extensive experiments on public ensembles demonstrate the efficiency of our algorithms, with MT-WAE computations in the orders of minutes on average. We show the utility of our contributions in two applications adapted from previous work on merge tree encoding (9). First, we apply MT-WAE to merge tree compression, by concisely representing them with their coordinates in the final layer of our auto-encoder. Second, we document an application to dimensionality reduction, by exploiting the latent space of our auto-encoder, for the visual analysis of ensemble data. We illustrate the versatility of our framework by introducing two penalty terms, to help preserve in the latent space both the Wasserstein distances between merge trees, as well as their clusters. In both applications, quantitative experiments assess the relevance of our framework. Finally, we provide a C++ implementation that can be used for reproducibility.