METRICS FOR PERSISTENCE DIAGRAMS: AN OPTIMAL TRANSPORT VIEW.

THÉO LACOMBE

Persistence diagrams (PDs) appear as a core tool to encode topological information in data analysis. PDs provide a concise way to summarize the underlying topology of a given object at all scales—informally as a locally finite point cloud supported on the upper half-plane $\{(t_1, t_2) \in \mathbb{R}^2, t_2 > t_1\}$. The space of these diagrams can be equipped with partial matching metrics, with theoretical guarantees on the stability of diagrams under pertubations of input data. However, the computational cost of such metrics is known to be prohibitive in large-scale applications, and the structure these metrics induce on the space of PDs is non-linear. This makes the use of standard statistical tools or machine learning techniques—even as simple as estimating Fréchet means or barycenters of a sample of PDs—challenging. We present a way to address these issues by reformulating PD metrics as an optimal *partial* transport problem [1], and show how recent advances in computational optimal transport [2, 3, 4] can be adapted to deal efficiently with large samples of PDs. In particular, regularizing the optimal transport problem with an entropic penalization yields a convex problem that can be solved efficiently with the Sinkhorn algorithm. Unlike previous methods to approximate PD metrics, this algorithm can be parallelized and implemented efficiently on GPUs. These approximations are also differentiable, leading to a simple and scalable method to estimate barycenters of PD samples. We showcase the strength of this approach by estimating the Fréchet means and performing k-mean clustering with diagram metrics on large PD samples [5].

References

[1] A.Figalli, N.Gigli: *A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary condition.* J. Math. Pures Appl.(9), 2010, p 107-130.
[2] M.Cuturi: *Sinkhorn Distances: Lightspeed computation of optimal transport.* Advances in neural information processing systems, 2013, p 2292-2300.
[3] M.Cuturi, A.Doucet: *Fast computation of Wasserstein barycenters.* International Conference of Machine Learning, 2014, p 685-693.
[4] G.Peyré, M.Cuturi: Computational optimal transport. URL: `http://arxiv.org/abs/1803.00567`
[5] T.Lacombe, M.Cuturi, S.Oudot: *Large Scale computation of Means and Clusters for Persistence Diagrams using Optimal Transport.* Advances in neural information processing systems, 2018.

(Théo Lacombe) Datashape - Inria Saclay
Email address: `theo.lacombe@inria.fr`

Grant from AMX, École polytechnique.