Barycentres de descripteurs topologiques via Transport Optimal

Théo LACOMBE, DataShape, Inria Saclay

Les diagrammes de persistance constituent un outil standard en analyse topologique des données : à un objet (nuage de points, graphe, complexe simplicial), on associe un diagramme de persistance qui en décrit la topologie (présence composantes connexes, boucle, cavités, etc.) à toutes les échelles. L'espace des diagrammes de persistance $\mathcal{D}$ peut être vu comme un espace de mesures localement finies et supportées sur le demi-plan $\{(t_1, t_2) \in \mathbb{R}^2,\ t_2 > t_1\}$. Cet espace peut être muni d'une métrique, notée d dans la suite, et permet de définir et d'étudier des descripteurs statistiques construits à partir des diagrammes de persistance.

Dans cet exposé, nous présenterons le problème d'estimation de barycentres (ou "moyennes de Fréchet") dans l'espace des diagrammes de persistance en utilisant un formalisme issu du transport optimal avec bord [1]. Formellement, étant donné une famille de diagrammes $\mu_1 \dots \mu_N$, le problème d'estimation d'un barycentre de $\mu_1 \dots \mu_N$ consiste à trouver (ou approcher) une solution de

$$\operatorname{argmin} \left\{ \frac{1}{N} \sum_{i=1}^{N} d(\nu, \mu_i)^p,\ \nu \in \mathcal{D} \right\},$$

où p est un paramètre ($p = 1$ donne une notion de médiane, $p = 2$ correspond à une moyenne arithmétique). Ce problème peut se généraliser à des familles infinies de diagrammes, il s'agit alors de considérer une distribution de probabilité $\mathbb{P}$ supportée sur $\mathcal{D}$, et de minimiser $\nu \mapsto \int d(\nu, \mu)^p \mathrm{d}\mathbb{P}(\mu)$. Nous donnerons un résultat d'existence et de "consistance" des barycentres dans ce cadre général, et évoquerons comment ce formalisme donne lieu à des algorithmes d'estimation rapide de barycentre en pratique, c'est-à-dire en considérant une famille finie de diagramme finis [2].

Références

[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] TL, 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
theo.lacombe@inria.fr