

Titre : Contributions statistiques à l'algorithme Mapper

Mot clés : Analyse Topologique de Données, Statistique, Apprentissage Automatique

Résumé : Depuis son introduction comme une méthode basée sur la topologie pour effectuer de la visualisation de données et de l'inférence, l'algorithme Mapper a été utilisé avec succès dans de nombreuses applications. Cette thèse fournit une analyse statistique du Mapper suivant trois points de vue : l'optimisation, l'approximation normale et le transport optimal.

Premièrement, nous donnons une méthode d'optimisation d'hyperparamètres qui permet de choisir parmi une famille paramétrique la fonction filtre, qui est cruciale pour la construction du graphe Mapper. L'optimisation est basée sur un critère topologique et permet au Mapper de s'approcher des méthodes de visualisation classiques ne nécessitant pas de

grands ajustements.

Deuxièmement, nous établissons un théorème central limite pour des fonctions d'homologie persistante calculées sur le graphe Mapper. Nous utilisons des résultats d'approximation normale dans les espaces de Poisson et des outils de géométrie stochastique portant sur la connectivité des graphes géométriques aléatoires.

Finalement, nous analysons le Mapper à travers la géométrie des espaces métriques mesurés et prouvons sa convergence vers son graphe de Reeb correspondant en distance de Gromov-Wasserstein. Nous prouvons aussi la stabilité des graphes de Reeb en ce qui concerne les perturbations de la fonction filtre avec cette métrique.

Title: Statistical contributions to the Mapper algorithm

Keywords: Topological Data Analysis, Statistics, Machine Learning

Abstract: Since its introduction as a topology based method for data visualization and inference, the Mapper algorithm has been successfully used in numerous applications. This thesis provides a statistical analysis of the Mapper following three points of view: optimization, normal approximation and optimal transport.

First, we provide a hyperparameter optimization scheme that allows to choose from a parametric family the filter function that is crucial to the construction of the Mapper graph. The optimization is based on a topological criteria and allows to make the Mapper more in line with classical off-the-shelf data visualiza-

tion methods.

Second, we establish a central limit theorem for persistent homology functions that are computed on the Mapper graph. We use normal approximation results for Poisson spaces and stochastic geometry tools for the connectivity of random geometric graphs.

Finally, we use a metric measure space geometry outlook to analyze the Mapper and prove its convergence to its continuous Reeb graph counterpart in the Gromov-Wasserstein distance. We also prove the stability of Reeb graphs to filter perturbations with respect to this metric.