Statistical analysis and parameter selection for Mapper

Bertrand Michel, Ecole Centrale de Nantes - Laboratoire Jean Leray bertrand.michel@ec-nantes.fr

The Mapper algorithm is a method for topological data analysis by Gurjeet Singh, Facundo Mmoli and Gunnar Carlsson. In this work, we study the question of the statistical convergence of the 1-dimensional Mapper to its continuous analogue, the Reeb graph. We show that the Mapper is an optimal estimator of the Reeb graph, which gives, as a byproduct, a method to automatically tune its parameters and compute confidence regions on its topological features, such as its loops and flares. This allows to circumvent the issue of testing a large grid of parameters and keeping the most stable ones in the brute-force setting, which is widely used in visualization, clustering and feature selection with the Mapper.

Joint work with: Mathieu Carriere (INRIA DataShape) and Steve Oudot (INRIA DataShape).

References

- [1] M. Carriere, B. Michel and S. Oudot Statistical analysis and parameter selection for Mapper. arXiv preprint arXiv:1706.00204, 2017.
- [2] Gurjeet Singh, Facundo Mmoli and Gunnar Carlsson Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition Eurographics Symposium on Point Based Graphics, European Association for Computer Graphics, , pp. 91?100, 2007.