

Thesis: Piecewise Linear Manifold Clustering

Abstract: This work studies the application of topological analysis to non-linear manifold clustering. A novel method, that exploits the data clustering structure, allows to generate a topological representation of the point dataset. An analysis of topological construction under different simulated conditions is performed to explore the capabilities and limitations of the method, and demonstrated statistically significant improvements in performance. Furthermore, we introduce a new information-theoretical validation measure for clustering, that exploits geometrical properties of clusters to estimate clustering compressibility, for evaluation of the clustering goodness-of-fit without any prior information about true class assignments. We show how the new validation measure, when used as regularization criteria, allows creation of clusters that are more informative. A final contribution is a new meta-clustering technique that allows to create a model-based clustering beyond point and linear shaped structures. Driven by topological structure and our information-theoretical criteria, this technique provides structured view of the data on new comprehensive and interpretation level. Improvements of our clustering approach are demonstrated on a variety of synthetic and real datasets, including image and climatological data.

Committee:

- Professor Robert M. Haralick, mentor, The Graduate Center
- Professor Mikel Vejdemo-Johansson, College of Staten Island
- Professor Michael Grossberg, The City College of New York

Outside Member:

- Juri Katz, S&P Capital IQ