Rainer von Sachs

(Louvain-la-neuve, Belgien)

Title:

"AdaCLV for Interpretable Variable Clustering and Dimension Reduction of Spectroscopic Data"

Abstract:

Nuclear Magnetic Resonance (NMR) spectroscopy is a popular method for measuring samples containing an unknown mixture of molecules. In NMR data, the observed spectrum for a given mixture of K molecules is essentially the sum of K unobserved pure spectral profiles plus noise. Each molecule has its own pure spectral profile, or "signature," which contains one or more peaks whose height is proportional to the molecule's concentration. For certain applications, such as biomarker discovery in medicine, it may be useful to identify the spectral regions where peaks occur and to cluster these spectral variables into groups, with one cluster for each molecule. Sparse regression or classification techniques, such as the Group-Lasso, can then be used to identify regions of the spectrum that most predict a given health outcome. In this presentation, we propose a novel method, AdaCLV, that performs variable clustering, dimension reduction and estimates the most prominent signatures of the molecules measured by NMR. We compare AdaCLV with several competitor methods and show that it is less sensitive to tuning parameters, and it estimates molecular signatures and latent variables as well or better than the other methods.