STABILITY SELECTION FOR OMICS DATA

Ron Wehrens and Pietro Franceschi

Biostatistics and Data Management, Fondazione Edmund Mach
San Michele all'Adige (TN)
Italy
http://cri.fmach.it/BDM

The recent fields of transcriptomics, metabolomics, proteomics, often summarized under the heading "omics", aim at providing holistic views of biological systems: by measuring as many variables as possible it is hoped that relevant biological information can be uncovered. In many cases, this comes down to investigating which variables are important when comparing classes, when following samples over time or when relating omics measurements with phenotypical observations. Given the extremely low sample-to-variable ratios in typical omics data sets, standard biomarker identification methods like t tests tend to work not very well. Multiple testing corrections are indispensible but often lead to an unacceptable loss of power.

Stability selection (Meinshausen and Buehlmann, 2010) provides a way to avoid many of the false positives in biomarker identification by repeatedly subsampling the data, and only considering those variables as putative biomarkers that consistently show up as important. They used the lasso as a primary variable selection method. In our own work, we have shown that also selecting the largest coefficients in non-sparse regression models such as PLS works well, when combined with the stability selection framework (Wehrens et al. 2011). We support these claims with the analysis of several experimental and simulated data sets. In particular, the BioMark package for R, implementing stability selection as well as Higher Criticism thresholding, contains an experimental spike-in data set from the area of metabolomics, which can aid in further algorithm testing and development. From these analyses, it follows that stability selection is a very general and robust framework for variable selection.

References:
Meinshausen N, Buehlmann P (2010). "Stability selection." J. R. Statist. Soc. B, 72, 417 473. With discussion.
Wehrens R, Franceschi P, Vrhovsek U, Mattivi F (2011). "Stability-based biomarker selection." Anal. Chim. Acta, 705, 15 23.