Better multiple Testing: Using multivariate co-data for hypotheses

Daniel Fridljand
European Molecular Biology Laboratory, Heidelberg;
Institute of Applied Mathematics, University of Heidelberg

Consider a multiple testing task, where for each test we have access to its p-value and additional information represented by a uni- or multivariate covariate that may be informative on prior probabilities of null and alternative hypotheses and/or on the test's power. The independent hypothesis weighting (IHW, Ignatiadis and Huber, 2021) framework uses these covariates to stratify the tests into a finite number of bins which are then assigned different weights for use in a weighted multiple testing method such as that of Benjamini and Hochberg. IHW guarantees false discovery rate control and increases overall power (number of discoveries) compared to the unweighted method.

(Ignatiadis and Huber, 2021) propose performing the stratification by quantile slicing the covariates. However, this simple approach does not take full advantage of the data. First, quantiles are not able to capture the heterogeneities among tests. The hypotheses within each bin should be as homogeneous as possible. Second, the procedure becomes computationally infeasible for high-dimensional covariates, due to an exponential increase of the number of bins with the covariates' dimension.

We address this challenge by introducing a random forest based approach, where the leaves of the trees replace the bins. The objective function is chosen such that the splits are sensitive to the prior probability and shape of the postulated conditional density. This yields homogeneous bins and hence increases power. The trees can handle high-dimensional covariates and an orthonormal series expansion makes growing the trees computationally efficient. We apply the updated implementation to RNA-seq data and show the benefits of our method when performing differential expression analysis.

Ignatiadis, N. and Huber, W. (2021) Covariate powered cross-weighted multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **83**, 720–751. John Wiley and Sons Inc. DOI: 10.1111/rssb.12411.