

Resampling-Based Analysis of Microarray Data

Stan Young
GlaxoSmithKline
5 Moore Drive RTP NC 27709 USA
919 483 8456
919 315 4175
genetree@bellsouth.net
CAMDA Datasets 1 and 2

Young S. Stanley, Emptage Michael, Yount Eric, Westfall Peter

Resampling-based analysis of micro-array data S. Stanley Young¹, Michael Emptage¹, Eric Yount², Peter Westfall³ ¹Glaxo Wellcome, ²NCSU, ³Texas Tech University

Researchers are starting to collect micro-array data with the hopes of relating some biological event, e.g. cancer, to gene expression. Often there are a large number of genes under consideration, thousands, and the number of training samples is low, tens to a few hundred. There is great interest in isolating a few genes that are predictive and in having a statistical measure of the strength of the relationship. Our idea is to use resampling-based methods to isolate relationships and measure their strength. Resampling can take into account multiple testing corrected for correlation and distributional characteristics. The methods are applied to the Golob et al. data set.

Keywords

Resampling, multiple testing

Tools

The statistical analysis was done using SAS Institute proc multtest. Example code is given on the poster. A description of the technology is given in the book, Resampling-Based Multiple Testing, Wiley, by Peter Westfall and Stan Young