BMC Systems Biology

Poster presentation

Variance stabilising transformations for NMR metabolomics data

Helen Parsons* and Mark Viant

Address: School of Biosciences, The University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Email: Helen Parsons* - hmp166@bham.ac.uk

* Corresponding author

This abstract is available from: http://www.biomedcentral.com/1752-0509/1?issue=S1

Classifying the fingerprint of an NMR spectrum is a crucial step in many metabolomics experiments. Since many classification techniques such as principal component analysis (PCA) depend upon variance discrepancies, it is important to first maximise any contribution from wanted class variance between biological samples and minimise any contribution from unwanted technical variance arising from the preparation of the samples and measurement of the NMR metabolic fingerprints. The generalised logarithm (glog) transform was developed to stabilise the variance between technical replicates in a two component error model [1] and has also been applied to NMR spectra previously [2]. To increase the effectiveness of the transform on NMR spectra, the glog was extended to include a baseline offset term. This decreases the unwanted noise contribution on the transformed spectra. The extended glog transformation is given as:

$$z = \ln\left((y - y_0) + \sqrt{(y - y_0)^2 + \lambda} \right)$$

for z the transformed intensity and y the original intensity of the spectra. y_0 and λ are transformation parameters which are to be found.

Here we have applied the extended glog transform to technical replicates of NMR spectra of tissue extracts from marine mussels, to determine the optimised transformation parameters λ and y_0. Next we applied the optimised transformation to a data set comprised of two classes of NMR spectra from stressed and unstressed mussels. Following transformation, the results show significantly better separation of the classes on a PCA scores plot than can be achieved with both untransformed data and also data transformed using Pareto scaling, a widely used method in NMR metabolomics [3]. In conclusion, we have demonstrated the value of the extended glog transformation to stabilise the technical variance in an NMR metabolomics dataset and have achieved significantly improved classification of NMR fingerprints from stressed and unstressed animals.

References

1. Rocke D, Lorenzato S: **A two-component model of measurement error in analytical chemistry.** *Technometrics* 1995, **37(2):**176-184.
2. Purohit P, Rocke D, Viant M, Woodruff D: **Discrimination models using variance-stabilizing transformation of metabolomic NMR data.** *OMICS* 2004, **8(2):**118-130.
3. Keun H, Ebbels T, Antti H, Bollard M, Beckonert O, Holmes E, Lindon J, Nicholson J: **Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling.** *Analytica Chimica Acta* 2003, **490(1):**265-276.