Biclustering: Clustering (High Dimensional) Two-Way Data

Sebastian Kaiser

Institutskolloquium
10.12.2008

Abstract. Over the last decade, bicluster methods have become more and more popular in different fields of two way data analysis and a wide variety of algorithms and analysis methods have been published. After Cheng and Church (2000) followed the initial bicluster idea of Hartigan (1972) and started to calculate bicluster on microarray data, a wide range of different articles were published dealing with different kinds of algorithms and methods to preprocess and analyze the results of such methods. Comparisons of several bicluster algorithms can be found, e.g., in Madeira and Oliveira (2004) or Prelic et al. (2006).

The focus of this talk is to give a short overview on the general theory of biclustering, to give some theoretical background of the most popular methods and their algorithms, and to introduce the R package `biclust`, which contains a collection of bicluster algorithms, preprocessing methods for two way data, and validation and visualization techniques for bicluster results.

References

CHENG, Y. and CHURCH, G. M. (2000): Biclustering of expression data. In: *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*, 1,93–103.

HARTIGAN, J.A. (1972): Direct Clustering of a data matrix. *Journal of The American Statistical Association, 67,12079–12084.*

MADEIRA, S. C. and OLIVEIRA, A. L. (2004): Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 1(1),24–45.*

PRELIC, A., BLEULER, S., ZIMMERMANN, P., WIL, A., BÜHLMANN, P., GRUISSEM, W., HENNING, L., THIELE, L., and ZITZLER, E. (2006): A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics, 22(9),1122–1129.*