Leibniz
Universität Hannover

Partner in der Forschung

Institut für Technische Chemie, Callinstr. 3, 30167 Hannover

Clustering - a powerful tool for discovering new correlations in genomics

Christine Klockow, Frank Stahl, Bernd Hitzmann
Leibniz University of Hannover, Institute of Technical Chemistry

Introduction

Over the last years, a fast development in the area of microarrays and the analysis of gene expression can be witnessed. Complete genomes and their expression profiles can be examined in comparatively short time intervalls, resulting in an impressive mass of information. However, it is still a great challenge to interpret the biological meaning of the data. One possible approach is to cluster the expression profiles into groups of genes with similar behavior.

K-means clustering- The algorithm

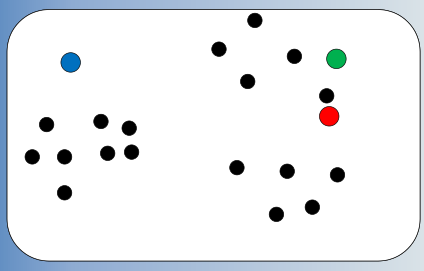

Fig. 1: Three centroids (blue, red & green) are randomly added to the data points (black)

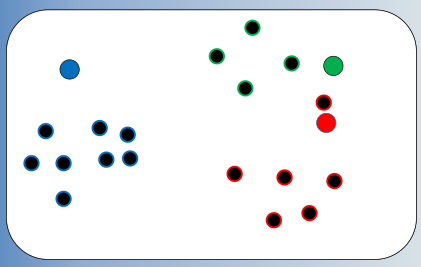

Fig. 2: Data points are assigned to nearest centroid

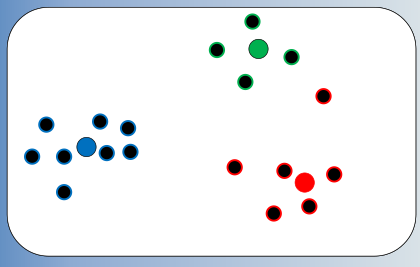

Fig. 3: Centroids are shifted into center of their data points

Step 2 and 3 are repeated until a good separation is achieved.

Results

The algorithm was applied to a data set derived from a microarray experiment with yeast cells grown at different glucose concentrations. The set contained about 4000 genes, which were sorted into 20 clusters.

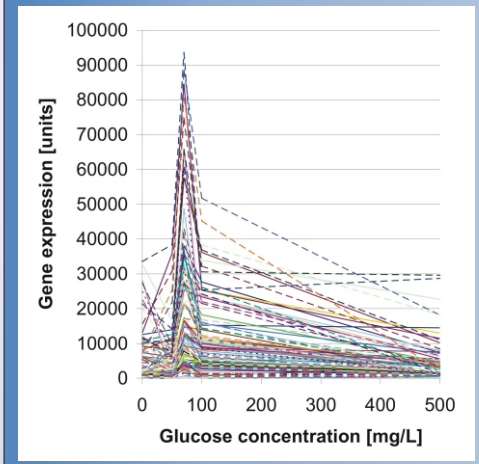

Fig.4: Expression profile of cluster 1 generated by k-means clustering. Each line represents one gene.

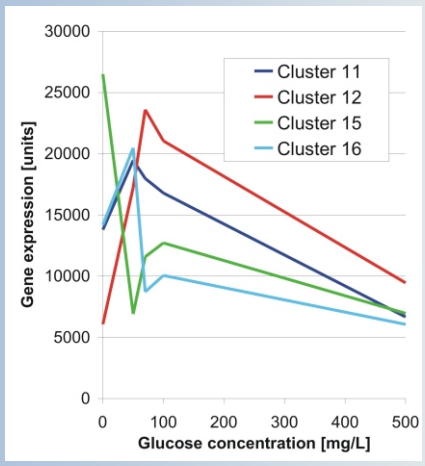

Fig. 5: Mean values for three clusters generated by k-means clustering

Figure 4 shows the expression profile of one of the resulting clusters. Mainly genes from glycolysis, cell cycle and cell division are present in this cluster. The clusters displayed in Fig. 5 contain many genes from the peptide metabolism.

Conclusions

The k-means clustering did not only yield the examplary results presented above, but many more valueable hints about gene regulation. In other clusters, genes from the categories catabolism, stress response, ketone or vitamin metabolism assembled.
Also, facts already known - like stress response during starvation - could be reconstructed by this method, which demonstrates that it renders reliable results.