

Gaussian Mixture Based Method for Missing Value Estimation in Microarray Data

Background DNA microarray technology is extensively used for the generation of gene expression data from different cell lines and disease tissues. The expression levels of thousands of genes can be assayed in a single chip. Clustering is routinely used in the analysis of microarray data, whereby genes with similar expression profiles can be identified. Clustering could also be used to classify disease tissues. The distance among entities is measured using measures such as the Pearson correlation coefficient, the Spearman rank correlation coefficient, and inner product. However, missing entries arising from blemishes on the microarray often complicate the measurement of the distances. Thus, it is critical to determine the values of missing entries in performing cluster analysis.

Description of the Technology

UMDNJ researchers have designed a method to estimate the value of missing entries based on Gaussian Mixture modeling. This method consists of clustering microarray data using a Gaussian mixture clustering model and estimating missing values by a GMCimpute algorithm.

Advantages

This method has been found to be empirically more accurate than existing methods such as K nearest neighbor method, the singular value decomposition method, and the spine fitting method.

Applications

- To estimate the values of missing entries in microarray data.

Deliverables

- Computer program

Patent Status

- PCT patent application filed. (Application No.: PCT/US2004/024351)

Licensing Opportunity

- This technology is available for non-exclusive license.

Contact

Peter Golikov, MS, MBA
Director, Ventures and Licensing
University of Medicine and Dentistry of New Jersey
335 George Street
New Brunswick, NJ 08901

Direct Phone: (732)-235-9355
Main Office Phone: (732)-235-9350
Facsimile: (732)-235-9358
golikope@umdnj.edu

File RWJ 03-14 /Welsh