

gustav / May 27, 2011 10:15PM

[\[Interdisciplinary\]\[Information\]\[Medicine\] NCKU Landmark Project: Construction of an Intelligent Disease Biomarker Mining Platform by Integrated Analysis of Clinical, Genomic and Proteomic Databases](#)[Interdisciplinary][Information][Medicine] NCKU Landmark Project: Construction of an Intelligent Disease Biomarker Mining Platform by Integrated Analysis of Clinical, Genomic and Proteomic Databases ([Chinese Version](#))

NCKU Research Express (2011/05/27) The cDNA microarray technology has enabled the experimental study of thousands of gene expressions simultaneously to investigate different kinds of biological processes. In general, microarray data can be expressed in the form of a large matrix, where each row indicates a gene expression vector under a number of different experimental conditions (samples). The conditions may correspond to different time points, different cells or different environmental conditions.

Recently, various kinds of analysis methods were proposed to extract biological knowledge for different study purposes, like clustering analysis, association rule analysis, classification analysis, etc. Traditional microarray data clustering methods, such as hierarchical clustering and k-means clustering techniques, are used to group genes according to their similarity of expression profiles in whole conditions. However, the biologically relevant genes usually have similar expression profile only under a subset of conditions. Therefore, a number of biclustering techniques have been proposed for identifying a subset of genes sharing compatible expression patterns.

Among the existing biclustering methods, however, few biclustering algorithms provide a query based (query-driven) approach for biologists to search the bicluster that contains a certain gene of interest, e.g. a potential drug target. LIU et al. (2007) proposed a query-driven biclustering method for a user-defined gene, called MSBE, by using a new similarity score between two genes and similarity scores for a bicluster. However, how to set the suitable cut threshold for parameters is a major problem in MSBE since a strict cut threshold leads to loss of important genes or conditions in finding a bicluster.

In the NCKU Landmark Project (CHEN et al. 2010, Vincent Shin-Mu TSENG as the principal investigator), the team extended the MSBE algorithm to propose a novel query-driven biclustering algorithm, namely Weighted Fuzzy-based Maximum Similarity Biclustering (WF-MSB) with a generalized fuzzy approach for a user-defined (reference) gene g^* . In contrast to MSBE method, each difference value between g^* and other genes under each condition is converted to a fuzzy set with a set of fuzzy regions R ($|R| = v$). The different v degrees of expression closeness, between g^* and other genes, are defined by v membership functions corresponding to different fuzzy regions. Based on this fuzzy approach, these v biclusters for various closeness degrees can be detected to represent different types of similarity degrees to the reference gene g^* , e.g. the most similar bicluster or the most dissimilar bicluster to the reference gene. In particular, the most dissimilar bicluster could provide a new viewpoint to discover a set of genes with opposite view to the given reference gene in gene expression analysis, e.g. the apoptosis-related genes versus the anti-apoptosis-related genes, the repressor of transcription factors versus the target genes and etc. The team points out, this is the first consideration for biclustering on gene expression data. Moreover, different from most of traditional biclustering, a more significant bicluster, with higher expression values, can be detected by applying weighting approach to these conditions in which the reference g^* have more significant expression levels.

Further Information:

[NCKU Research Express 2011/05/27](#)

Edited 2 time(s). Last edit at 05/27/2011 10:17PM by gustav.