

## **EFFICIENT GENERALIZED MATRIX APPROXIMATIONS FOR BIOMARKER DISCOVERY AND VISUALIZATION IN GENE EXPRESSION DATA**

Wenyuan Li, Yanxiong Peng, Hung-Chung Huang and Ying Liu  
Department of Computer Science, University of Texas at Dallas  
Richardson, TX 75083, U.S.A.

In most real-life gene expression data sets, there are often multiple sample classes with ordinals, which are categorized into the normal or diseased type. The traditional feature or attribute selection methods consider multiple classes equally without paying attention to the up/down regulation across the normal and diseased types of classes, while the specific gene selection methods particularly consider the differential expressions across the normal and diseased, but ignore the existence of multiple classes. In this paper, for improving the biomarker discovery, we propose to make the best use of these two aspects: the differential expressions (that can be viewed as the domain knowledge of gene expression data) and the multiple classes (that can be viewed as a kind of data set characteristic). Therefore, we simultaneously take into account these two aspects by employing the 1-rank generalized matrix approximations (GMA). Our results show that the consideration of both aspects can not only improve the accuracy of classifying the samples, but also provide a visualization method to effectively analyze the gene expression data on both genes and samples. Based on the GMA mechanism, we further propose an algorithm for obtaining the compact biomarker by reducing the redundancy.