

An integrative gene selection with association analysis for microarray data classification

ABSTRACT

The rising interest in integrative approach has shifted gene selection from purely data-centric to incorporating additional biological knowledge. Integrative gene selection is viewed as a promising approach in microarray data classification that took into consideration the complex relationships among genes. However, in most of the existing methods, the selection of genes is still based on expression values alone and biological knowledge is integrated at the end of analysis to verify experimental results or to gain biological insights. Thus, this paper proposed an integrative gene selection based on filter method and association analysis for selecting genes that are not only differentially expressed but also informative for classification. Association analysis is employed to integrate microarray data with multiple types of biological knowledge simultaneously, and to identify groups of genes that are frequently co-occurred in target samples. It has been tested on four cancer-related datasets, and two types of biological knowledge are incorporated, namely Gene Ontology (GO) and KEGG Pathways (KEGG). The experimental results show that the recommended GO based models, KEGG based models, and GO-KEGG based models outperformed the expression-only models by attaining better classification accuracies with lesser number of genes. The performance of the integrative models verified the efficiency and scalability of association analysis in mining microarray data.

Keyword: Association analysis; Classification; Gene selection; Integrative; Microarray data