

Improving biomarker discovering for chemosensitivity prediction using an integrated approach

A. Visconti¹, F. Cordero¹, R.A. Calogero²✉

¹Department of Computer Science, University of Torino, Torino, Italy

²Molecular Biotechnology Center, University of Torino, Torino, Italy

Motivations

Recently, high-throughput techniques have been successfully used to investigate different aspects of the cell behaviour, opening new perspectives in fields such as molecular biology, medicine, and pharmacology. These advancements have led to the arising of new areas of research such as pharmacogenomics: the discipline studying the influence of genetic variations on drug response having the objective of optimizing drug therapies so to ensure the maximum of efficacy and the minimum of side effects. A crucial step in pharmacogenomics is the discovery of genes (biomarkers) that are responsible for drug responses. By using each gene as a 'feature' and the drug response as the 'class' to be predicted, the problem can be casted as the one of 'finding the set of features that allows the best prediction of the class'. This is a known machine learning task known as the 'feature selection problem'. Feature selection algorithms are usually classified into two categories: filter and wrapper approaches. In literature, several filter methods have been described for the task of detecting biomarkers. However, the kind of statistical analysis carried on by filter approaches cannot capture interactions among genes. Wrapper methods use the prediction performances of a given machine learning approach to assess the usefulness of a subset of genes. They consider reciprocity among genes and obtain remarkable performances. Nevertheless, the information about genes interactions in biological pathways is still missed.

Methods

In this work we propose an integrated approach to detect biomarkers liable for cell lines respons-

es to drug administration. Specifically, our approach integrates: i) a filter and a wrapper technique for biomarker discovering, and ii) different sources of knowledge, namely transcriptional profiles, drugs activity, and pathways interactions. The proposed approach is composed by two steps. In the first step, we apply a filter method to identify a set candidate biomarkers. This is a pre-processing method exploiting a priori knowledge about gene expression levels and gene interactions. In detail, we identify differential expressed genes using the Rank Product methodology. Then, we perform a pathway analysis for extracting genes (i.e. network hubs and bottlenecks) that are likely to be responsible for the measured differential gene expression levels. In the second step, we utilize a wrapper approach to single out biomarkers. To this purpose, we use multiple runs of a genetic algorithm to assess the importance of each candidate biomarker.

Results

We use the NCI60 DNA panel to test our approach. It consists in an in vitro screening of several chemical compounds over 60 human cancer cell-lines. We select 118 drugs whose mechanisms of action are known. To assess the quality of the proposed approach, we compare the obtained accuracies to those of a rough wrapper technique, namely Random Forests. Random Forests has been considered standard 'tool-box of methods' for class prediction and gene selection with microarray data. We outperform Random Forests approach. Finally, we analyze the extracted biomarkers by using the 'Ingenuity systems' showing that they are strictly related to the targets of administrated drugs.