

Michał Zientek

SILESIAAN UNIVERSITY OF TECHNOLOGY. FACULTY OF AUTOMATIC CONTROL,
ELECTRONICS AND COMPUTER SCIENCE. INSTITUTE OF COMPUTER SCIENCE
e-mail: michal.zientek@gmail.com

Paweł Foszner

SILESIAAN UNIVERSITY OF TECHNOLOGY. FACULTY OF AUTOMATIC CONTROL,
ELECTRONICS AND COMPUTER SCIENCE. INSTITUTE OF COMPUTER SCIENCE

Andrzej Polański

SILESIAAN UNIVERSITY OF TECHNOLOGY. FACULTY OF AUTOMATIC CONTROL,
ELECTRONICS AND COMPUTER SCIENCE. INSTITUTE OF COMPUTER SCIENCE

**Improving functional coherence of gene signatures by using
Gene Ontology terms**

Molecular classifiers based on gene expression profiles obtained in DNA microarray experiments are very extensively studied due to their potential to be applied in a variety of areas, such as diagnosis, prediction of therapy results etc. The specific property of classification of gene expression profiles is the importance of the feature selection step. This stems from the fact that in DNA microarray experiments very large numbers of values of genes expressions are obtained for relatively small number of samples.

Therefore in recent years significant effort has been paid to development of feature selection algorithms leading to choosing appropriate subsets of genes, called gene signatures, which are then used as arguments for discriminant function in the molecular classifier.

Among methods for gene selection, proposed in the literature, an interesting group are algorithms using the idea of combining the information on expressions of genes with the information on functional coherence of the set of selected genes. Several papers in the literature showed that such an approach can lead to improvement in classification quality.

In our study we propose an algorithm based on the Steiner tree metrics, which was recently proposed as a tool for measuring functional coherence of subsets of genes. The proposed method uses a recursive procedure for signature slimming by removing least coherent genes. The obtained signature has largest measures of functional coherence. We present the use of the proposed algorithm for classification of several publicly available DNA microarray datasets.

This work was financially supported by the Polish Ministry of Science under Grant No. N516 441938 Efficient methods of genome browsing based on the Burrows Wheeler Transform.