Abstract

Microarray is a rich topic which gives the opportunity for researchers to classify cancer samples without any previous biological knowledge. Microarrays high dimensionality
characteristic motivated the importance of gene selection techniques. In this paper a new filter multiple scoring gene selection technique MGS-CM is proposed. This technique is further combined with three classifiers to introduce three new classification systems (MGS-SVM, MGS-KNN and MGS-LDA) which are validated and evaluated on three microarray datasets. The proposed MGS-CM technique was proven to be an efficient technique as it extracts the highly informative genes reducing the original datasets by at least 99.6%. Also two of the three proposed classification systems guaranteed the perfect classification (100%) of the leukemia microarray samples. The third one classifies the lymphoma microarray samples with only two misclassifications which is the minimum recorded number. The proposed systems achieved very good results and guaranteed reliable classification for new unclassified samples.

References

- S.-B. CHO, H.-H. WON, "Data Mining for Gene Expression Profiles from DNA Microarray", International Journal of Software Engineering and Knowledge Engineering,
MGS-CM: A Multiple Scoring Gene Selection Technique for Cancer Classification using Microarrays

- Online]. Available: http://citeseer.ist.psu.edu/contextsummary/1048862/0

Index Terms

Computer Science

Bioinformatics
Keywords

Cancer Classification  Microarrays
Multiple scoring gene selection