

Abstract

Gene expression analysis is the use of quantitative RNA measurements of gene expression in order to characterize biological processes and clarify the mechanisms of gene transcription. The microarray technology provides a great help for gene expression analysis; however, there are several general issues in using microarray data. The datasets are usually complex and noisy. Another important issue is high dimensionality by typically having fewer than one hundred instances where each instance quantifies the expression levels of several thousands of genes. As a result, traditional data mining methods cannot be effectively applied to gene expression classification and there is a need for more dedicated techniques to approach this problem. As one solution, feature construction and feature selection have been applied to preprocess the gene expression data in a way to overcome the high dimensionality problem.

In this talk I will first present common feature selection techniques and then a novel approach for identifying relevant genes by employing a fuzzy classifier. First a fuzzy classifier rule set is derived such that each rule involves a compact set of genes. Then, a correlation matrix is produced by considering the correlations between the genes in each rule. Frequent pattern mining is applied on the correlation matrix to find the maximal sets of correlated genes after tuning the minimum support value. Experiments conducted on the Leukemia and Prostate datasets demonstrate the effectiveness of the proposed approach in producing relevant genes.

References:

F. Zeng, C. H. R. Yap and L. Wong: Using Feature Generation and Feature Selection for Accurate Prediction of Translation Initiation Sites," *Genome Informatics*, Vol.13, pp.192-200, 2002.

M. A. Hall: Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning," *Proc. of ICML*, pp.359-366, 2000.

H. Ishibuchi, T. Nakashima and T. Muratam: Performance Evaluation of Fuzzy Classifier Systems for Multi-dimensional Pattern Classification Problems," *IEEE Trans. on Systems, Man, and Cybernetics*, October 1999.

Li M. Fu and E. S. Youn: Improving reliability of gene selection from microarray Functional Genomics data," *IEEE Transactions on Information Technology in Biomedicine*, Vol.7, No.3, 2003.

Khabbaz, M., Kianmehr, K., Alshalalfa, M. and Alhaji, R. (2007) Fuzzy Classifier Based Feature Reduction for Better Gene Selection. *Proceedings of the International Conference on Data Warehouseing and Knowledge Discovery*, Springer-Verlag LNCS, Germany.