Abstract

Diagnosis of cancer is one of the most emerging clinical applications in microarray gene expression data. However, cancer classification on microarray gene expression data still remains a difficult problem. The main reason for this is the significantly large number of genes present relatively compared to the number of available training samples. In this paper, a novel approach to feature extraction combining the statistical t-test and absolute scoring is proposed for achieving better classification rate. Suitable classification approaches using the linear Support Vector Machines, the Proximal Support Vector Machines and the Newton Support Vector Machines is also discussed. A comparative analysis on the different techniques for feature extraction is also presented. Microarray cancer data based on Adenoma and Carcinoma with 7086 and 7457 genes of 4 and 18 patients respectively is used for this study. Increase in the classification rate of the proposed new method is clearly demonstrated in the results.

References

- J. P. Florido, H. Pomares, I. Rojas, J. M. Urquiza, L. J. Herrera, M. G. Claros,
Two Step Feature Extraction Method for Microarray Cancer Data using Support Vector Machines

Wei Du, Yan Wang, De-Ping Wang, Zhong-Bo Cao, Ying Sun and Yan-Chun Liang.

"Effect of Pre-processing methods on Microarray-based SVM classifiers in Affymetrix Genechips"; International Joint Conference on Neural Networks (IJCNN), pp 1-6, 2010

Azadeh Mohammadi, Mohammad Hossein Sarae, De-Ping Wang, Zhong-Bo Cao, Ying Sun and Yan-Chun Liang.

"An Improved Normalized Signal to Noise Ratio Method for Irrelevant Genes Removing"; 3rd International Conference on Biomedical Engineering and Informatics (BMEI 2010), pp 2275-2279, 2010

Nicholas A. Furlotte, Lijing Xu, Robert W. Williams, Ramin Homayouni.


Jinn-Yi Yeh, Tai-Shi Wu, Min-Che Wu, Der-Ming Chang.

"Applying Data Mining Techniques for Cancer Classification from Gene Expression Data"; International Conference on Convergence Information Technology, IEEE Computer Society, pp 703-708, 2007

Huang, D.; Chow, T. W. S.; Ma, E. W. M.; Jinyan Li.


Tang, Yuchun; Zhang, Yan-Qing; Huang, Zhen.


Guang-bin Huang; Hongming Zhou; Xiaojian Ding; Rui Zhang.


Glenn Fung & O. L. Mangasarian.

"Finite Newton Method for Lagrangian Support Vector Machine Classification";

Xiong Fu-song.


Ruopeng Wang; Hongmin Xu; Hong Shi.


IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol 8, Issue 1, pp 94-107, 2011

Alireza Osareh, Bita Shadgar.

"Microarray Data Analysis for Cancer Classification"; 5th International Symposium on Health Informatics and Bioinformatics,
Turkey, pp 125-132, April 20-22, 2010
- Chen Liao, Shutao Li, Zhiyuan Luo, "Gene Selection for Cancer Classification using Wilcoxon Rank Sum Test and Support Vector Machine", International Conference on Computational Intelligence, pp 368-373, 2006
- Yuchun Tang, Yan-Qing Zhang, and Zhen Huang, "Development of Two-Stage SVM-RFE Gene Selection Strategy for Microarray Expression Data Analysis", IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, VOL. 4, NO. 3, Published by the IEEE CS, CI, and EMB Societies & the ACM, pp 365-381, July-September 2007
- Patharawut Saengsiri, Sageemas Na Wichian, Phayung Meesad, Unger Herwig, "Comparison of Hybrid Feature Selection Models on Gene Expression Data", Eighth International Conference on ICT and Knowledge Engineering, pp 13-18, 2010
- Shutao Li, Chen Liao, James T. Kwok, "Wavelet-Based Feature Extraction for Microarray Data Classification", International Joint Conference on Neural Networks, Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada, pp 5028-5033, July 16-21, 2006

Index Terms
Computer Science
Artificial Intelligence
Keywords
Normalization  Linear SVM  Proximal SVM  Newton SVM  absolute scoring