Memetic Computation Based SVM for Cancer Classification

Nga, Nguyen Thi、吳幸珍

E-mail: 360072@mail.dyu.edu.tw

ABSTRACT

Cancer is one of the dreadful diseases found in most of the living being, which is one of the challenging studies for scientist towards 21th century. In cancer diagnosis and treatment, cancer classification plays a very important role. With the advent of DNA microarrays technology, constructing gene expression profiles for different cancer types has already become a promising means for cancer classification. However, it offers a challenge for current machine learning research. Microarray datasets are characterized by high dimension and small sample size. Over-fitting is a major problem due to the high dimension, while the small data size makes it worse. Support vector machine (SVM) is statistical classification algorithm that classifies data by separating two classes with the help of a functional hyper plane. SVM is known for good performance on noisy and high dimensional data such as microarray. One main disadvantage of using SVMs is that the performance of classifier depends on setting of parameters. In this thesis, we do classify cancer using gene expression data with a SVM classifier. A hybrid approach of particle swarm optimization (PSO) and simulated annealing (SA) is proposed to determine proper setting of SVM parameters which can improve the quality of SVM model. Our approach is a combination of methods. The motivation is to bring out an effective classification method for cancer by utilizing the strength of various techniques and compensating for their weaknesses. The proposed approach is tested on six benchmark cancer gene expression data sets, namely, colon, leukemia, lung, ovarian, prostate and breast. The experimental results show that the classification accuracy rates of the proposed method are competitive to that of other existing methods. It can be used as an efficient computational tool for microarray data analysis.

Keywords : cancer classification, support vector machine, parameter optimization.

Table of Contents

REFERENCES

[1] M. Eisen and P. Brown, “ DNA Arrays for Analysis of Gene Expression ” , Methods Enzymology, vol. 303, pp. 179-205, 1999.
[2] R. Lipshutz, S. Fodor, T. Gingeras, and D. Lockhart, “ High Density Synthetic Oligonucleotide Arrays ” , Nature Genetics, vol. 21, pp. 20-24, 1999.
[3] G. Sophia Reena, P. Rajeswari, “ A Survey of Human Cancer Classification using Micro Array Data ” , Int. J. Comp. Tech. Appl., vol. 2 (5), pp. 1523-1533, 2011.
[4] F. Chu, L. Wang, “ Application of Support Vector Machines to Cancer Classification with Microarray Data ” , International Journal of Neural Systems, vol. 15, no. 6, pp. 475 – 484, 2005.
[5] S. Wang, J. Wang, H. Chen and B. Zhang, “ SVM-Based Tumor Classification with Gene Expression Data ” , ADMA 2006, LNAI 4093, pp. 864-870.
[6] J. Phan, R. Moffitt, J. Dale, J. Petros, A. Young, M. Wang, “ Improvement of SVM Algorithm for Microarray Analysis Using Intelligent Parameter Selection ” , Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, 2005.
[7] O. Okun, H. Priisalu, “ Ensembles of Nearest Neighbors for Gene Expression Based Cancer Classification ” , Studies in Computational Intelligence (SCI) 126, pp. 115 – 134, 2008.
[8] R. Diaz-Uriarte, S. Alvarez-de Andres, “ Gene Selection and Classification of Microarray Data using Random Forest ” , BMC Bioinformatics 7: 3, 2006.
[9] A. Hasan, “ Evaluation of Decision Tree Classifiers and Boosting Algorithm for Classifying High Dimensional Cancer Datasets ” , International Journal of Modeling and Optimization, vol. 2, no. 2, April 2012.
[10] J. Khan, J. Wei, M. Ringner, L. Saal, M. Ladanyi, “ Classification and Diagnostic Prediction of Cancers using Expression Profiling and Artificial Neural Networks ” , Nat Med 7, pp. 673 – 679, 2001. 29 [11] P. Rajeswari and G. Sophia Reena, “ Human Liver Cancer Classification using Microarray Gene Expression Data ” , International Journal of Computer Applications (0975 – 8887), vol. 34, no. 6, November 2011.
[12] Lee JW, Lee JB, Park SH M Song, “ An Extensive Comparison of Recent Classification Tools Applied to Microarray Data ” , Computational Statistics and Data Analysis 2005, 48:869 – 885.
[13] A. Statnikov, C.F. Aliferis, I. Tsamardinos, D. Hardin, S. Levy, “ A Comprehensive Evaluation of Multicategory Classification Methods for Microarray Gene Expression Cancer Diagnosis ” , Bioinformatics 21 (5), pp. 631 – 643, 2005.
[14] V. N. Vapnik, The Nature of Statistical Learning Theory, 2nd ed., New York: Springer-Verlag, 1999.
[15] B. Scholkopf, C. J. C. Burges and A. Smola, Advances in Kernel Methods: Support Vector Learning, Cambridge, MA: MIT Press, 1999.
[16] K.M. Chung, W.C. Kao, C.L. Sun, L.L. Wang, C.J. Lin, “ Radius Margin Bounds for Support Vector Machines with RBF Kernel ” , Neural Comput. 15 (11), November 2003.
[17] F. Friedrichs, C. Igel, “ Evolutionary Tuning of Multiple SVM Parameters ” , Neurocomputing 2004, 64:107 – 17.
[18] V. N. Vapnik. The Nature of Statistical Learning Theory. NewYork: Springer-Verlag, 1995.
[19] J. Kennedy, RC. Eberhart, “ Particle swarm optimization ” , In:Proc IEEE International Conference on Neural Networks, IEEEservice Center, Piscataway, N.J., 4:1942 – 1948, 1995.
[20] M. A. M. de Oca, T. Stutzle, M. Birattari, and M. Dorigo, “ Frankenstein ’ s PSO: A Composite Particle Swarm Optimization Algorithm ” , IEEE Transactions on Evolutionary Computation, vol. 13, no. 5, October 2009. 30 [21] C-C. Chang and C-J. Lin. LIBSVM: A Library for Support Vector Machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[22] H. Hu, J. Li, A. Plank, H. Wang, G. Daggard, “ A Comparative Study of Classification Methods for Microarray Data Analysis ” , Proc. Fifth Australasian Data Mining Conference (AusDM), 2006.
[23] L. Guo-Zheng, Z. Xue-Qiang, Y. Y. Jack, Y. Mary Qu, “ Partial Least Squares Based Dimension Reduction with Gene Selection for Tumor Classification ” , Bioinformatics and Bioengineering (BIBE), 2007.
[24] P. Yanxiong, L. Wenyuan, and L. Ying, “ A Hybrid Approach for Biomarker Discovery from Microarray Gene Expression Data for Cancer Classification ” , Cancer Informatics, pp. 301 – 311, 2006.
[25] Y. Jieping, L. Tao, X. Tao, and J. Ravi, “ Using Uncorrelated Discriminant Analysis for Tissue Classification with Gene Expression Data ” , IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 1, no. 4, Oct-Dec 2004.