Available online at www.sciencedirect.com
ScienceDirect
Applied Soft Computing 8 (2008) 1664–1669
Applied Soft Computing
www.elsevier.com/locate/asoc

A novel ensemble of classifiers for microarray data classification

Yuehui Chen *, Yaou Zhao

Computational Intelligence Laboratory, School of Information science and Engineering, University of Jinan, 106 Jiwei Road, 250022 Jinan, PR China

Received 8 January 2007; received in revised form 10 January 2008; accepted 15 January 2008
Available online 20 January 2008

Abstract

Micorarray data are often extremely asymmetric in dimensionality, such as thousands or even tens of thousands of genes and a few hundreds of samples. Such extreme asymmetry between the dimensionality of genes and samples presents several challenges to conventional clustering and classification methods. In this paper, a novel ensemble method is proposed. Firstly, in order to extract useful features and reduce dimensionality, different feature selection methods such as correlation analysis, Fisher-ratio is used to form different feature subsets. Then a pool of candidate base classifiers is generated to learn the subsets which are re-sampling from the different feature subsets with PSO (Particle Swarm Optimization) algorithm. At last, appropriate classifiers are selected to construct the classification committee using EDAs (Estimation of Distribution Algorithms). Experiments show that the proposed method produces the best recognition rates on four benchmark databases.

Keywords: Microarray classification; Estimation of distribution algorithms (EDA); Particle swarm optimization (PSO); Ensemble learning; Correlation analysis; Fisher-ratio

1. Introduction

Microarray technology has provided the ability to measure the expression levels of thousands of genes simultaneously in a single experiment. Each spot on a microarray chip contains the clone of a gene from a tissue sample. Some mRNA samples are labelled with two different kinds of dyes, for example, Cy5 (red) and Cy3 (blue). After mRNA interact with the genes, i.e., hybridization, the color of each spot on the chip will change. The resulted image reflects the characteristics of the tissue at the molecular level [1].

In recent years, research has showed that accurate cancer diagnosis can be achieved by performing microarray data classification. Various intelligent methods have been applied in this area. But the microarray data consists of a few hundreds of samples and thousands or even ten thousands of genes. It is extremely difficult to work in such a high dimension space using traditional classification methods directly. So gene selection methods have been proposed and developed to reduce the dimensionality. These include principal components analysis (PCA) [8], Fisher-ratio, *t*-test, and correlation analysis. Along with the feature selection methods, intelligent methods have been applied for microarray classification, such as support vector machine (SVM) [7], *K* nearest neighbor (KNN) [6], artificial neural network (ANN) [1]. But high accurate classification is difficult to achieve. Most intelligent classifiers are apt to be over-fitted. Recent years, ensemble approaches have been proposed. It combines multiple classifiers together as a committee to make more appropriate decisions for classifying microarray data instances. It offers improved accuracy and reliability. Much research has shown that a sufficient and necessary condition the approach outperforms its individual members is that the base classifiers should be accurate and diverse. An accurate classifier is one that has an error rate of better than randomly guessing classes for new instances, and two classifiers are diverse if they make different errors on common data instances [9]. So there are two important aspects to be focused on ensemble approaches.

First aspect is how to generate diverse base classifiers. In traditional, re-sampling has been widely used to generating training datasets for base classifiers learning. This method is much too random and due to the small numbers of samples, the datasets may be greatly similar. In this paper, different methods such as correlation analysis, Fisher-ratio are firstly applied to generate feature subsets. Since variety of selection methods, some features selected in the subsets are different with each other and all of them are informative genes. Then re-sampling

* Corresponding author.
E-mail addresses: yhchen@ujn.edu.cn (Y. Chen), yaou_zhao@yahoo.com.cn (Y. Zhao).

1568-4946/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2008.01.006

the feature subsets to form learning datasets. Owing to the datasets forming from different feature subsets, it may be much more various and more efficient.

The second aspect is how to combine the base classifiers. In this paper, an intelligent approach for constructing ensemble classifiers is proposed. The methods first training the base classifiers with particle swarm optimization (PSO) algorithm, and then select the appropriate classifiers to construct a high performance classification committee with estimation of distribution (EDA) algorithm. Experiment show that the proposed methods produce the best recognition rates.

The paper is organized as follows. The feature selection methods are introduced in Section 2. The particle swarm optimization used to train the neural networks which are employed as the base classifiers is depicts in Section 3. The optimal design method for constructing ensemble classifiers is described in Section 4. Section 5 gives the simulation results. Finally, we present some concluding remarks.

2. Gene selection methods

Although there are a large number of genes in microarray, only small parts of genes have great impact on classification. Lots of genes are always similar in cancer and normal cases. Even worse, some genes may act as ''noise'' and undermine the classification accuracy. Hence, to obtain good classification accuracy, we need to pick out the genes that benefit the classification most. In addition, reducing the number of genes can help to cut down the inputs for computation, so the classifiers are much more efficient for classification and run much faster. Recently, many techniques are proposed to select the informative genes and two methods employed in this paper are introduced as follows.

2.1. Feature selection based on correlation analysis

In order to score the similarity of each gene, two ideal feature markers are defined. The two ideal feature markers are negatively correlated to represent two different aspects of classification boundaries. The feature markers are a binary vectors consisting of 0 and 1. The first feature marker is 1 in class A and 0 in class B, and the second is 0 in class A and 1 in class B. The two markers are expressed as follows:

$$\text{ideal}_1 = (1,1,1,1,0,0,0,0,0,0)$$
$$\text{ideal}_2 = (0,0,0,0,1,1,1,1,1,1)$$

The two feature makers are highly correlated to classes. If the genes are similar with the markers (the distance from the marker and the gene is small.), we consider that the genes are informative for classification. For calculate the distance of each gene, four measures are used in this paper:

Pearson Correlation (PC)

$$\text{PC} = \frac{\sum_{i=1}^{n}(\text{ideal}_i - \mu_{\text{ideal}})(g_i - \mu_g)}{\sqrt{\sum_{i=1}^{n}(\text{ideal}_i - \mu_{\text{ideal}})^2}\sqrt{\sum_{i=1}^{n}(g_i - \mu_g)^2}} \quad (1)$$

Spearman Correlation (SC)

$$\text{SC} = 1 - \frac{6\sum_{i=1}^{n}(\text{ideal}_i - g_i)^2}{n \times (n^2 - 1)} \quad (2)$$

Euclidean Distance (ED)

$$\text{ED} = \sqrt{\sum_{i=1}^{n}(\text{ideal}_i - g_i)^2} \quad (3)$$

Cosine Coefficient (CC)

$$\text{CC} = \frac{\sum_{i=1}^{n}\text{ideal}_i \times g_i}{\sqrt{\sum_{i=1}^{n}\text{ideal}_i^2 \times \sum_{i=1}^{n} g_i^2}} \quad (4)$$

where n is the number of samples; μ_g is the mean of the gene and μ_{ideal} is the mean of ideal marker; g_i is the ith real value of the gene vector and ideal_i is the correspond ith binary value of the ideal marker vector.

The feature selection steps are given as follows:

(1) Use the measures to score all the genes in the data. (The score is good if the measurement is small.)
(2) Choose the first $N/2$ best score genes for ideal marker one and rest for ideal marker two. (N is the total number of features which are selected.)

After the two steps, the genes are ranked in terms of their significance and the feature subsets are generated according to the different measures.

2.2. Fisher-ratio for gene selection

Fisher-ratio [10] is a ratio between-class distances to with-in class distances. If there are two classes (e.g. cancer vs. normal) in a dataset, each sample is labeled with $Y \in \{+1, -1\}$ and gene express vector i is defined as $x_i = \{x_1^i, \ldots, x_n^i\}, 1 \leqq i \leqq m$, where m is the number of samples. For each gene i, the standard deviation σ_i^+ (resp., σ_i^-) and the mean μ_i^+ (resp., μ_i^-) are calculated and the Fisher-ratio F_i is:

$$F_i = \frac{(\mu_i^+ - \mu_i^-)^2}{\sigma_i^{+2} + \sigma_i^{-2}} \quad (5)$$

Gene with highest F_i value is most informative and the expression levels differ most on average in the two classes while also favoring those with small deviation in the respective classes. Then the genes with high F_i values are selected as the top features.

2.3. Generating the training set for ensemble classifiers

In tradition, two basic re-sampling methods have been widely used in generating the training dataset which are the re-sampling of instances (samples) and sub-sampling of features, but the two methods all have their shortcomings in processing microarray data. The microarray data are often extremely asymmetric in dimensionality, such as thousands or even tens of thousands of

genes and a few hundreds of samples. If the sub-sampling of features method is used for generating the training sets without feature selection, the genes which are chosen are much too random and we can not guarantee that the genes are most informative and efficient. There may be many useless genes left in the training set which would disturb the classification accuracy. If we only use the re-sampling of instances method to form the training sets, it is also not a good idea because every sample has too many genes to be learned by any classifier.

In this paper, the two methods are combined to generate the training sets. Firstly, different feature selection methods have been applied to get the genes which are more efficient for classification. The feature subsets generated from Pearson Correlation, Cosine coefficient, Euclidean Distance, Spearman Correlation and Fisher-ratio are denoted as S_{PC}, S_{CC}, S_{ED}, S_{SC}, S_{FR}. Secondly, we randomly select a set S ($S \in \{S_{PC}, S_{CC}, S_{SC}, S_{ED}, S_{FR}\}$) as a candidate feature subset. Thirdly, the training samples are randomly sampling from the set S to form new training data sets.

After k-run of sampling, a pool of training sets is produced for training base classifiers. Since variety of feature subsets, the training sets are different with each other. More importance, all of them have informative genes in different aspects. It may be much more various and efficient for classification.

3. Learning the datasets with neural networks

There are many kinds of methods for classification. In recent years, most researchers applied the SVM (Support Vector Machine) as a classifier to learn the microarray dataset and obtained very good results. But the SVM is very complex to compute and training the SVM costs a lot of time. If many SVMs are used as base classifiers for assembling, the training time may be very long and the ensemble classifiers are inefficient. Moreover, the SVMs are similar due to their learning algorithms, the ensemble classifiers can not increase the accuracy of classification efficiently. So selecting SVM as base classifier is not a good choice. In this paper, we use the artificial neural networks as the base classifiers and train them with PSO algorithm. The artificial neural network is simple to learn and can easily construct different explicit classification functions by changing the number of hidden-layer nodes. PSO is a global optimization algorithm, so the parameters of neural networks can gain very good values.

3.1. Artificial neural networks

Funahashi [15] has shown that neural networks with at least one hidden layer can approximate a variety of conditions. The neural networks typically consists of three neural layers: an input layer, a hidden layer and an output layer. All the neurons in one layer are connected with all the neurons in the next layer. In this type of network, the input layer is determined by the incoming signals. This upper layer distributes the input signals to neurons in the hidden layer. Each hidden neuron sums all its input signals by a dot product between its input vector and its weight and then adds a "bias" input. The final result is then transformed by an "activation function" (in neural network terminology) to produce an input signal to the output layer. The output layer processes its input signals in the same fashion. The entire process can be written mathematically as

$$y_k = f_o(\beta_k + \sum_j \omega_{jk} f_h(\beta_j + \sum_i \omega_{ij} x_i)) \quad (6)$$

where x_i is the input signal, y_k is the output signal, ω_{ij} is the weight between input neuron i to hidden neuron j, and ω_{jk} is the weight between hidden neuron j to output neuron k. The β_j and β_k are the biases for the hidden and output layers, and f_h and f_o are activation functions for the hidden and output layers. The logistic function defined as

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (7)$$

3.2. Parameter optimization with PSO

The Particle Swarm Optimization conducts searches using a population of particles which correspond to individuals in evolutionary algorithm (EA). A population of particles is randomly generated initially. Each particle represents a potential solution and has a position represented by a position vector x_i. A swarm of particles moves through the problem space, with the moving velocity of each particle represented by a velocity vector v_i. At each time step, a function f_i representing a quality measure is calculated by using x_i as input. Each particle keeps track of its own best position, which is associated with the best fitness it has achieved so far in a vector p_i. Furthermore, the best position among all the particles obtained so far in the population is kept track of as p_g. In addition to this global version, another version of PSO keeps track of the best position among all the topological neighbors of a particle. At each time step t, by using the individual best position, p_i, and the global best position, $p_g(t)$, a new velocity for particle i is updated by

$$v_i(t+1) = v_i(t) + c_1\phi_1(p_i(t) - x_i(t)) + c_2\phi_2(p_g(t) - x_i(t)) \quad (8)$$

where c_1 and c_2 are positive constant and ϕ_1 and ϕ_2 are uniformly distributed random number in [0, 1]. The term v_i is limited to the range of $\pm v_{max}$. If the velocity violates this limit, it is set to its proper limit. Changing velocity this way enables the particle i to search around its individual best position, p_i, and global best position, p_g. Based on the updated velocities, each particle changes its position according to the following equation:

$$x_i(t+1) = x_i(t) + v_i(t+1). \quad (9)$$

4. Optimal design method for constructing ensemble classifiers

Select many classifiers for constructing the committee are better than all [2]. So we should select appropriate classifiers to

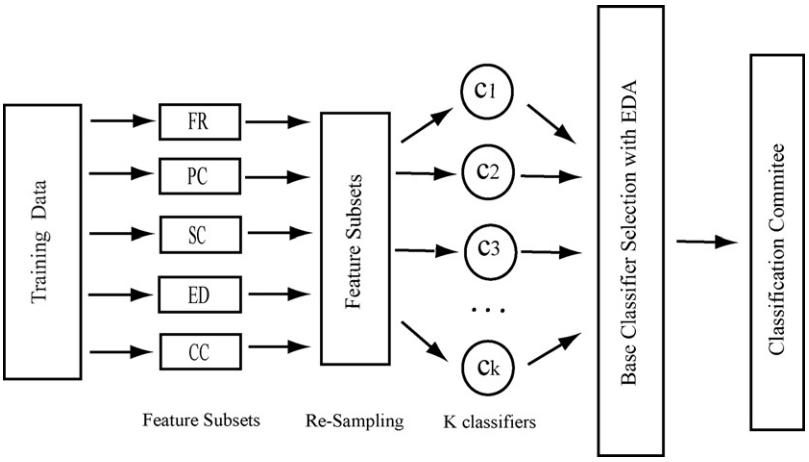

Fig. 1. FR, PC, SC, ED and CC indicate the feature sets which are generated by the feature selection methods (Fisher-ratio, Pearson Correlation, Cosine coefficient, Euclidean Distance and Spearman Correlation) respectively; $C_1, C_2, C_3, \ldots, C_k$ denote K base classifiers. The feature selection approaches are firstly employed to reduce the dimensionality of training data and a pool of candidate base classifiers is generated to learn the subsets which re-sampling from the different feature subsets with PSO algorithm. Then appropriate classifiers are selected to construct the classification committee using EDA.

form the classification committee. In traditional, many approaches can accomplish this task, such as greedy hill climbing. It evaluates all the possible local changes to the current set, such as adding one classifier to the set or removing one. It chooses the best or simply the first change that improves the performance of subset. Once a change is made for a subset, it is never reconsidered. But generally, it can not find the optimist solution. In this paper, we introduce a selection method using EDA algorithms.

4.1. Estimation of distribution algorithms

The EDA was first introduced by Larranaga and Lozano [3]. It is a search method that eliminates crossover and mutation from the Genetic Algorithm (GA) and places more emphasis on the relation between gene loci. More precisely, it generates the next generation based on probability distribution of N superior population samples. In this way, the probability distribution estimated at each generation is progressively converted into a probability distribution that generates more superior individuals. The EDA algorithm is given follows:

Step 0: Randomly generate a set of λ individuals ($t = 0$).
Step 1: Evaluate the λ individuals.
While (not done)
Step 2: Select μ individuals (where $\mu \leq \lambda$) to be parents. Develop a probability distribution/density function pt based on the parents.
Step 3: Create λ offspring using pt.
Step 4: Evaluate the offspring.
Step 5: The λ offspring replace the μ parents ($t = t + 1$).
Step 6: Goto While.

4.2. Constructing ensemble classifiers by EDA

Suppose K base classifiers are generated after trained by the feature subsets. They expressed as $C_1, C_2, C_3, \ldots, C_k$. S is the subsets of $\{C_1, C_2, C_3, \ldots, C_k\}$. Binary vectors are introduced to denote S. If C_i is selected, the ith position of the vector is 1; while C_i is not selected, the ith position is 0. Binary vectors are used to be chromosome of individuals and they can be evolved by EDA algorithm. In order to measure individuals, the fitness function should be created. We first generate the validation set $V$and then calculate the error E_v of each individual on V. $1/E_v$ is the fitness. E_v is depicted as follows:

$$E_{vi} = \sum_{j=1}^{K} p_{ij} \times \text{classifier}_j \tag{10}$$

Here E_{vi} is the error of the ith individual. K is the total number of base classifiers. P_{ij} is the binary number of chromosome at the jth position. classifier_j is the error of the jth base classifier on V. The flowchart of our methods is in Fig. 1.

5. Experiments

We performed extensive experiments on four benchmark cancer datasets, namely the Leukemia, Colon, Ovarian and Lungcancer.

5.1. The Leukemia dataset

The leukemia dataset was taken from a collection of leukemia patient samples reported by Golub et al. [4]. This well-known dataset often serves as benchmark for microarray analysis methods. It contains measurements corresponding to acute lymphoblast leukemia (ALL) and acute myeloid leukemia (AML) samples from bone marrow and peripheral blood. The dataset consisted of 72 samples: 25 samples of AML, and 47 samples of ALL. Each sample is measured over 7129 genes.

5.2. The Colon dataset

The samples from colon dataset were taken from colon adencarcinoma specimens snap-frozen in liquid nitrogen within

20 min of removal from patients [5]. The microarray dataset consists of 22 normal and 40 tumor tissue samples. In this dataset, each sample contains 2000 genes.

5.3. The Ovarian dataset

The goal of this experiment is to identify proteomic patterns in serum that distinguish ovarian cancer from non-cancer. This study is significant to women who have a high risk of ovarian cancer due to family or personal history of cancer. The proteomic spectra were generated by mass spectroscopy and the data set provided here includes 91 controls (normal) and 162 ovarian cancers. The raw spectral data of each sample contains the relative amplitude of the intensity at each molecular mass/charge (M/Z) identity. There are total 15,154 M/Z identities [17].

5.4. The Lungcancer dataset

Classification between malignant pleural mesothelioma (MPM) and adenocarcinoma (ADCA) of the lung. There are 181 tissue samples (31 MPM and 150 ADCA). The training set contains 32 of them, 16 MPM and 16 ADCA. The rest 149 samples are used for testing. Each sample is described by 12,533 genes [16].

The normalization procedure is firstly used for preprocessing the raw data. Four steps were taken:

(1) If a value is greater than the floor 16,000 and smaller than the ceiling 100, this value is replaced by the ceiling/floor.
(2) Leaving out the genes with $(\max - \min) \leq 500$, here max and min refer to the maximum and minimum of the expression values of a gene, respectively.
(3) Carrying out logarithmic transformation with 10 as the base to all the expression values.
(4) For each gene i, subtract the mean measurement of the gene μ_i and divide by the standard deviation σ_i. After this transformation, the mean of each gene will be zero, and the standard deviation will be one.

Table 1
Parameters used for experimets

Common parameters for PSO	
M: population size	20
c_1, c_2: learning factor	2.0
$v_{\max}$: the max velocity	1.8
x_{up}: the upper boundary of x	3.0
x_{down}: the lower boundary of x	−3.0
ϕ_1, ϕ_2: uniform random number	$[0, 1]$
Common parameters for EDA	
λ: population size	20
μ: elite size	5
Common parameters for Neural Network	
N_{i}: the number of input layer	30
N_{h}: the number of hidden layer	6–10
N_{o}: the number of output layer	1
$f_{\text{h}}, f_{\text{o}}$: activate function	$\frac{1}{1+e^{-x}}$

Table 2
Relative works on Leukemia dataset

Author	Test accuracy (%)
Our method	98.6
PSO + ANN	86.1
SVM [11]	94.1
C4.5 [12]	81.94
Nero-fuzzy [13]	87.5
KNN [14]	72.64

After this steps, the feature selection method is then employed to form the feature subsets and re-sampling method is used to form the training and testing datasets. In our experiment, 10 training datasets are generated and 30 informative features of each sample are extracted for training the 10 base classifiers. The neural network is employed to be the classifier and we use particle swarm optimization to adjust the weights of each neural network. Then EDA was applied for selecting appropriate NNs to constructing the classification committee. Table 1 indicates the parameters used for experiments.

For classification, we conducted leave-one-out cross-validation (LOOCV) for each dataset for comparing with the other peoples' work. In LOOCV, one of all samples is evaluated as testing data while the others are used as training data. The training data is used to select informative features. This process is repeated 20 times to obtain the average results with experiments in total.

Tables 2–5 compare PSO–ANN ensemble with single PSO-ANN classifier. For all the test dataset, the PSO–ANN ensemble can give the best classification accuracy. 2–12% increments are observed when using the ensemble technique. It means that the classifiers trained from the datasets which generated from different feature subsets are more diverse and the classifier ensemble is more efficient and robust than a single PSO–ANN.

A comparison of different feature extraction methods and different classification methods for leukemia, Colon, Lungcancer and Ovarian datasets are shown in Tables 3–5. For leukemia dataset, our method produces the best classification

Table 3
Relative works on Colon dataset

Author	Test accuracy (%)
Our method	95.2
PSO + ANN	88.7
SVM [11]	90.3
C4.5 [12]	85.48
Nero-fuzzy [13]	93.55
KNN [14]	75.81

Table 4
Relevant works on Lungcancer dataset

Method	Test accuracy (%)
Our method	100
PSO + ANN	98.3
TSP [18]	98.3
k-TSP [18]	98.9
DT [18]	96.13
NB [18]	97.79
KNN [18]	98.34

Table 5
Relevant works on Ovarian dataset

Method	Test accuracy (%)
Our method	99.6
PSO + ANN	97.0
NB [19]	96.2
BKS [19]	97.0
DT [19]	97.8
ORA [19]	98.3

accuracy 98.6%, while the other methods produce 84.6–95.8%. For colon data, our method gets a high classification accuracy 95.2% compared with the existing methods. For Lungcaner data, our method produces 100% classification accuracy, and for ovarian data, our method also performs best result of 99.6% compared to that of the others (96.2–98.3%). All the data sets show that our proposed method produces higher classification accuracy than the other methods.

6. Conclusions

In this paper, a novel ensemble of classifiers based on correlation analysis is proposed for cancer classification. The leukemia and colon databases are used for conducting all the experiments. Gene features are first extracted by the correlation analysis technique which greatly reduces dimensionality as well as maintains the informative features. Then the EDA is employed to construct the classifier committee for classification. Compare the results with some advanced artificial techniques, the proposed method produces the best recognition rates.

Acknowledgments

This research was supported by the NSFC under grant No.60573065, the Key Subject Research Foundation of Shandong Province and the Natural Science Foundation of Shandong Province (grant Y2007G33).

References

[1] F. Chu, L. Wang, Appliations of support vector machines to cancer classification with microarray data, Int. J. Neural Syst. 15 (6) (2005) 475–484.
[2] Z.-H. Zhou, J. Wu, W. Tang, Ensembling neural networks: many could be better than all, Artif. Intell. 137 (1–2) (2002) 239–263.
[3] P. Larranaga, J.A. Lozano, Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation, Kluwer Academic Publishers, 2001.
[4] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. GaasenBeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Blomfield, E.S. Lander, Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, Science 286 (12) (1999) 531–537.
[5] U. Alon, N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack, A.J. Levine, Broad patterns of gene expresson revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide array, Proc. Natl. Acad. Sci. 96 (12) (1999) 6745–6750.
[6] M.B. Eisen, B.O. Brown, DNA arrays for analysis of gene expression, Methods Enzymol. 303 (1999) 179–205.
[7] S.-B. Cho, Exploring features and classifiers to classify gene expression profiles of acute leukemia, Int. J. Pattern Recogn. Artif. Intell. 16 (7) (2002) 1–13.
[8] C.A. Harrington, C. Rosenow, J. Retief, Monitoring gene expression using DNA microarrays, Curr. Opin. Microbiol. 3 (2000) 285–291.
[9] T.G. Dietterich, Ensemble methods in machine learning, in: Proceeding of the First International Workshop on Multiple Classifier Systems, 2000, 1–15.
[10] K.Z. Mao, RBF neural network center selection based on Fisher ratio class separability measure, IEEE Trans. Neural Networks 13 (2002) 1211–1217.
[11] T.S. Furey, N. Cristianini, N. Duffy, D.W. Bednarski, M. Schummer, D. Haussler, Support vector machine classification and validation of cancer tissue samples using microarray expression data, Bioinformatics 16 (2000) 906–914.
[12] L. Yu, H. Liu, Redundancy based feature selection for microarray data, Department of Computer Science and Engineering Arizona State University, Technical Report, 2004.
[13] Z.Y. Wang, V. Palade, Y. Xu, Nero-fuzzy Ensemble approach for microarray cancer gene expression data analysis, in: Proceedings of the 2006 International Symposium on Evolving Fuzzy Systems (IEEE), 7–9, (2006), pp. 241–246.
[14] T. Jirapech-Umpai, S. Aitken, Feature selection and classification for microarray data analysis: Evolutionary methods for identifying predictive genes, Bioinformatics 6 (2005) 168–174.
[15] K. Funahashi, On the approximate realization of continuous mapping by neural networks, Neural Networks 2 (1989) 183–332.
[16] G. Gordon, R. Jensen, L. Hsiao, S. Gullans, J. Blumenstock, S. Ramaswamy, et al., Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma, Cancer Res. 62 (17) (2002) 4963–4967.
[17] E. Petricoin, A. Ardekani, B. Hitt, P. Levine, V. Fusaro, S. Steinberg, et al., Use of proteomic patterns in serum to identify ovarian cancer, Lancet 359 (9306) (2002) 572–577.
[18] A.C. Tan, D.Q. Naiman, L. Xu, R.L. Winslow, D. Geman, Simple decision rules for classifying human cancers from gene expression profiles, Bioinformatics 21 (20) (2005) 3896–3904.
[19] J.-H. Hong, S.-B. Cho, The classification of cancer based on DNA microarray data that uses diverse ensemble genetic programming, Artif. Intell. Med. 36 (2006) 43–58.

Yuehui Chen was born in 1964. He received his B.Sc. degree in the Department of Mathematics (major in control theory) from the Shandong University of China in 1985, and Ph.D. degree in Department of Electrical Engineering and Computer Science from the Kumamoto University, Japan in 2001. During 2001–2003, he had worked as a Senior Researcher of the Memory-Tech Corporation at Tokyo. Since 2003 he has been a member at the Faculty of School of Information Science and Engineering in University of Jinan, where he is currently heads the Laboratory of Computational Intelligence. His research interests include evolutionary computation, neural networks, fuzzy systems, hybrid computational intelligence and their applications in time-series prediction, system identification and intelligent control. He is the author and co-author of more than 100 papers. Professor Yuehui Chen is a member of IEEE, the IEEE Systems, Man and Cybernetics Society and the Computational Intelligence Society. He is also a member of the editorial boards of several technical journals and a member of the program committee of several international conferences.

Yaou Zhao was born in 1982. He received the B.S. degree in computer science from Jinan University, Jinan, China, in 2005. Since 2005, he has been a master student in the School of Information Science and Engineering, Jinan University. His research interests include evolutionary neural network, bioinformatics and face reorganization.