Neurocomputing 71 (2008) 3275–3281

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A probabilistic multi-class strategy of one-vs.-rest support vector machines for cancer classification

Jin-Hyuk Hong, Sung-Bae Cho *

Department of Computer Science, Yonsei University, 134 Shinchon-dong, Sudaemoon-ku, Seoul 120-749, Republic of Korea

ARTICLE INFO

Available online 24 June 2008

Keywords:
Support vector machines
Multi-class cancer classification
Naïve Bayes classifiers
Fusion method

ABSTRACT

Support vector machines (SVMs), originally designed for binary classification, have been applied for multi-class classification with effective decomposition and reconstruction schemes. Decomposition schemes such as one-vs.-rest (OVR) and pair-wise partition a dataset into several subsets of two classes so as to produce multiple outputs that should be combined. Majority voting or winner-takes-all is a representative reconstruction scheme to combine those outputs, but it often causes some problems to consider tie-breaks and tune the weights of individual classifiers. In this paper, we propose a novel method in which SVMs are generated with the OVR scheme and probabilistically ordered by using the naïve Bayes classifiers (NBs). This method is able to break ties that frequently occur when working with multi-class classification systems with OVR SVMs. More specifically, we use the Pearson correlation to select informative genes and reduce the dimensionality of gene expression profiles when constructing the NBs. The proposed method has been validated on several popular multi-class cancer datasets and produced higher accuracy than conventional methods.

1. Introduction

In the literature, support vector machines (SVMs) have been used for multi-class classification that is an important task in pattern recognition applications like bioinformatics. Even though SVMs show excellent performance in many applications, it is required to formulate a multi-class SVM method since it is originally designed for binary classification [3]. There are some direct approaches to developing a multi-class extension of SVMs such as Vapnik or Crammer and Singer [6], which often lead to a complex optimization problem. Instead, many researchers prefer to use several binary SVMs, in which a pool of SVMs is constructed according to a decomposition scheme and the outputs of the classifiers are combined by a reconstruction scheme [10]. One-vs.-rest (OVR), pair-wise and complete codes are representative decomposition schemes to generate a pool of SVMs, while majority voting, winner-takes-all, and error-correcting coding are the reconstruction schemes to combine those multiple outputs.

Hsu and Lin [10] compared various schemes of OVR, pair-wise and directed acyclic graph SVM (DAGSVM), where one-vs.-rest and DAGSVM were more suitable for practical use than pair-wise. Bredensteiner and Bennett [3] combined the linear programming and quadratic programming based on SVM for multi-class problems, while Angulo et al. [1] proposed K classes-support vector classification-regression (K-SVCR). Sebald and Bucklew [22] proposed M-ary SVM that uses only $\lceil \log_2(K) \rceil$ SVMs (K: # of classes), Gestel et al. [8] used Bayesian decoding to compose least squares SVMs (LS-SVMs) that repeatedly infer the posterior multi-class probabilities. Crammer and Singer [6] adopted output codes for combining the outputs of multiple classifiers.

In bioinformatics, SVMs have been also widely applied to classify multiple types of cancers. Lee [14] applied the multi-category SVM (MC-SVM) to classify leukemia data and small round blue cell tumors data, and Koo et al. [12] proposed an analysis of variance decomposition using structured kernels. Ramaswamy et al. [21] used OVR and winner-takes-all scheme to classify the GCM cancer dataset, where Yeang et al. [25] compared the SVMs with k-nearest neighbors and weighted voting for the same dataset. Liu et al. [17] combined the genetic algorithm and pair-wise SVMs where predictive genes are automatically determined through iteration. Li et al. [15] conducted a comparative study of feature selection and multi-class classification methods including SVMs, where Statnikov et al. [23] performed a systematic and comprehensive evaluation of several major algorithms for multi-class classification.

This paper proposes a novel multi-class classification approach integrating SVMs and naïve Bayes classifiers (NBs), especially applying to multi-class cancer classification with gene expression profiles. Since SVMs might manage the high-dimensional data like gene expression profiles, the original training dataset is used to

* Corresponding author. Tel.: +82 2 2123 3877; fax: +82 2 365 2579.
E-mail addresses: hjinh@sclab.yonsei.ac.kr (J.-H. Hong), sbcho@cs.yonsei.ac.kr (S.-B. Cho).

0925-2312/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2008.04.033

learn SVMs based on the OVR scheme. NBs are designed with a dataset consisting of genes selected by a rank-based feature selection method using Pearson correlation, which organize the SVMs probabilistically with respect to the subsumption architecture. Therefore, OVR SVMs are sequentially evaluated according to the probability of the classes obtained by NBs. The proposed method has been validated on several representative gene expression datasets including the GCM cancer dataset [12]. In the experiment on GCM cancer dataset, the dataset is divided into training and test sets in the manner of previous works in order to compare with them. Other experiments are conducted as leave-one-out cross validation.

2. Background

2.1. Cancer classification with gene expression profiles

Micro-array technologies recently developed produce large volume of gene expression profiles and provide richer information on diseases. It simultaneously monitors the expression patterns of thousands of genes under a particular experimental environment. With the technology, many researchers have studied cancer classification using gene expression profiles to achieve more accurate predictions. Several machine learning techniques have been developed and applied to many clinical problems by constructing classifiers or predictive models from data, and yielded promising results [5].

Especially the classification of cancers from gene expression profiles has been actively investigated in bioinformatics. It commonly consists of feature selection and pattern classification as shown in Fig. 1. At first, feature selection chooses informative features useful to categorize a sample into predefined classes among lots of genes in data. Pattern classification is composed of learning a classifier with those features and categorizing samples with the classifier learned [5].

Gene expression profiles provide useful information to classify different forms of cancers, but the data also include useless information for the classification. Therefore, only relevant ones should be extracted from them. It is well known that the irrelevant or redundant data degrade the accuracy of classification, so constructing an appropriate gene subset might be essential to learn a good classifier. Moreover, it is important to find a small subset of genes sufficiently informative to distinguish cancers for diagnostic purposes [11].

2.2. Multi-class classification using SVMs

SVMs, well studied in the statistical learning theory, have been actively investigated in pattern classification and regression. SVMs map an input sample to a high-dimensional feature space and try to find an optimal hyperplane that minimizes the recognition error for training data using a non-linear transformation function [1]:

$$X : x = (x_1, \ldots, x_n) \rightarrow F : \Phi(x) = (\Phi_1(x), \ldots, \Phi_n(x)) \quad (1)$$

Let n be the number of training samples. For the sample x_i with the class-label $c_i \in \{1,-1\}$, the SVM calculates

$$f(x) = \sum_{i=1}^{n} \alpha_i c_i K(x, x_i) + b, \quad K(x, x_i) = \Phi(x)\Phi(x_i) \quad (2)$$

Coefficient α_i in Eq. (2) is non-zero when x_i is a support vector that composes the hyperplane; otherwise it is zero. The kernel function $K(x, x_i)$ can be easily computed by an inner product of the non-linear mapping function. Table 1 shows some kernel functions such as linear, polynomial, Gaussian, and sigmoid.

Since SVM is a basically binary classifier, a decomposition strategy for multi-class classification is required such as OVR, pair-wise or complete-code methods [6]. As a representative scheme, the OVR strategy trains M (# of classes) SVMs, where each

Table 1
Kernel functions of SVMs

Linear	Polynomial	Gaussian	Sigmoid
$(x \cdot x_i)$	$(x \cdot x_i+\gamma)^d$	$\exp(-\|x-x_i\|^2/2\sigma^2)$	$\tanh(x \cdot x_i+\gamma)$

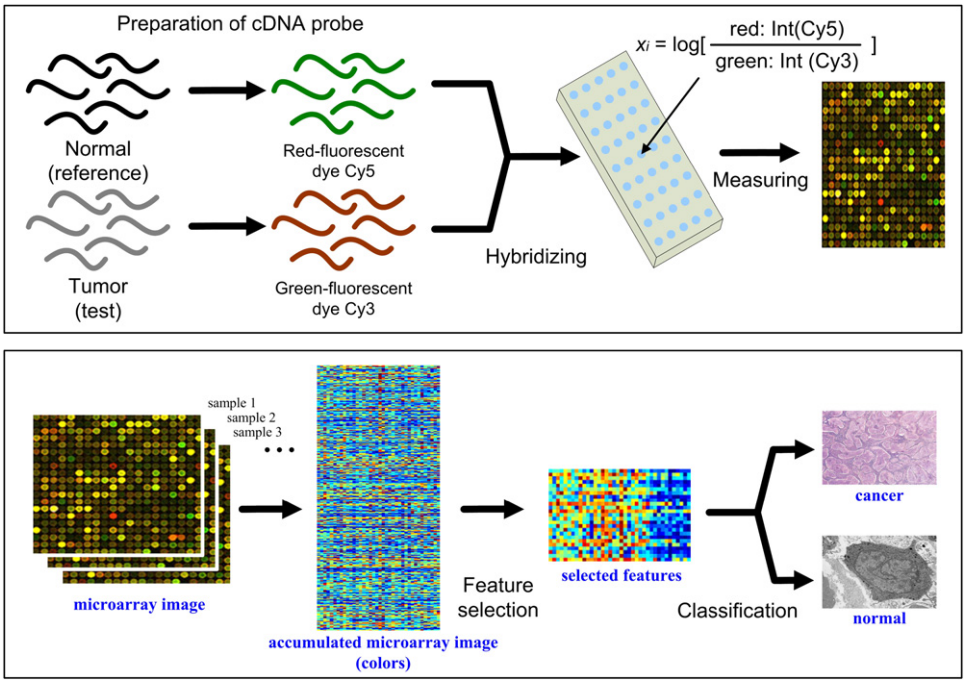


Fig. 1. Classification of gene expression profiles.

SVM classifies samples into the corresponding class against all the others. The decision function $f_j(x)$ of the jth SVM replaces c_i of Eq. (2) with t_i as follows:

$$t_i = \begin{cases} +1 \text{ if } c_i = j \\ -1 \text{ if } c_i \neq j \end{cases} \quad (3)$$

After constructing SVMs, a reconstruction strategy is required to combine multiple outputs of them. Popular methods for combination such as majority voting, winner-takes-all, error-correcting codes (ECCs), behavior knowledge space (BKS), and decision templates are widely employed.

- *Majority voting*: For a sample, this method simply counts the votes received from individual classifiers, and selects a class with the largest number of votes. An analytic justification may be given by the well-known Condorcet's theorem, while a theoretical study can be found in [13]. Although it is simple to achieve good performance, this method cannot handle cases where classifiers tie.
- *Winner-takes-all*: In order to resolve problems like ties caused by majority voting, this method classifies a sample into the class that receives the highest value among the L classifiers for the M-class problem. This is often known as maximum where $ind_{i,j}(x)$ is an indicator function with 1 if the label i is the positive class of the jth SVM, -1 if it is the negative class, and 0 otherwise.

$$c = \arg \max_{i=1,\ldots,M} \sum_{j=1}^{L} ind_{i,j}(x) d_j(x) \quad (4)$$

3. A hybrid classifier for multi-class cancer classification

Contrary to conventional methods having a static property, we propose a hybrid classifier that considers the probability of classes obtained by NBs to manage the ambiguity of OVR SVMs. Tie cases, which frequently occur when using OVR SVMs for multi-class classification, might decrease classification performance. The proposed method manages this possibility by organizing OVR SVMs based on the subsumption architecture. The subsumption architecture is a representative method used to select a proper action when there are multiple actions activated [4], while the order of OVR SVMs is determined by NBs in this paper.

The proposed method consists of NBs and OVR SVMs as shown in Fig. 2. NBs estimate the posterior probability for classes $prob = \{p_1, p_2, \ldots, p_m\}$ by using the training dataset that only includes informative genes, while SVMs classify samples by using the original training dataset according to the OVR scheme as explained in Section 2. The margin of a sample $o\text{-}svm = \{ma_1, ma_2, \ldots, ma_m\}$ is produced by OVR SVMs. In order to manage ambiguity in cases of ties (multiple SVMs satisfy) and rejections (no SVM satisfies), in this paper, the proposed method sequentially selects OVR SVMs, where the evaluation order is determined by the posterior probability of each class that NBs produces. The corresponding OVR SVM of a more probable class takes the precedence in subsumption architecture over the others.

When classifying a sample, the method first estimates the probability of each class by using NBs, and then organizes OVR SVMs as the subsumption architecture according to the probability. Finally, a sample is evaluated sequentially until an OVR SVM is satisfied. When an OVR SVM is satisfied, the sample is classified into the corresponding class, while it is classified into the class of the highest probability when no OVR SVMs are satisfied. Fig. 3 shows the pseudo code of the proposed method. Each OVR SVM has the corresponding threshold that is determined as a value among the outputs of the SVM for the training data, which produces the highest classification rate.

DNA micro-array data include the expression information of thousands or even tens of thousands of genes. Since it is hard to design NBs that include all the genes, a subset of informative genes is selected by using the feature selection process based on Pearson correlation. Cutting down the number of features to a sufficient minimum is often useful to improve classification performance [5].

We define two ideal markers that represent a standard of good features, and utilize the features by scoring the respective similarity with each ideal marker as shown in Fig. 4. Two ideal markers are negatively correlated to represent two different aspects of classification boundaries. The first marker is high in class A and low in class B, and the second marker is low in class A and high in class B. The first marker is a binary vector which consists of 1 for all the samples in class A and 0 for all the samples in class B, while the second marker is another binary vector which is composed of 0 for all the samples in class A and 1 for all the samples in class B. Since this feature selection method is originally designed for binary classification, we select features based on the OVR scheme. Ten genes are selected for each class: the first 5 for

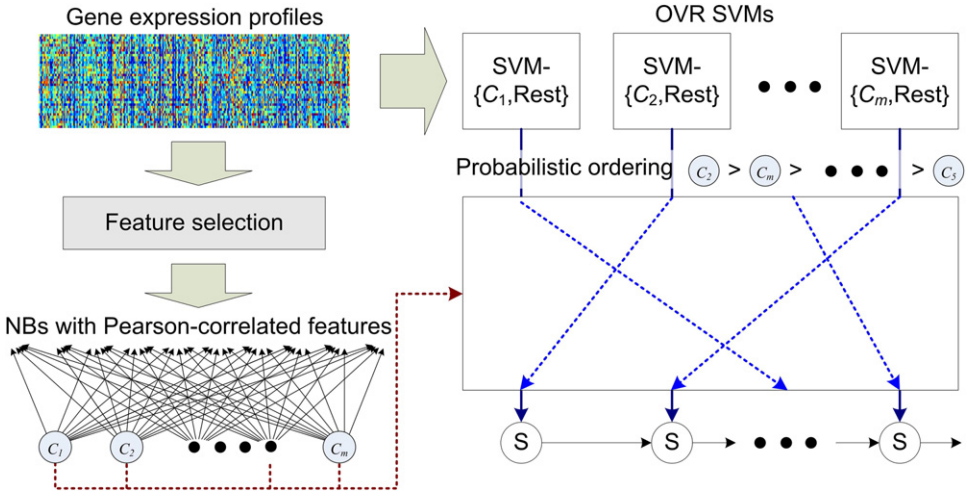


Fig. 2. Overview of the proposed method.

the ideal marker 1 and the rest for the ideal marker 2. When there are m classes, total $m \times 10$ genes are used to construct NBs.

The similarity between an ideal marker *ideal* and a gene g can be regarded as a distance, while the distance represents how far they are located from one another. A gene is regarded as an informative gene if the distance is small, while the gene is regarded as an uncorrelated gene if the distance is large. Pearson correlation is used to measure the similarity as follows:

$$PC = \frac{\sum_{i=1}^{n}(ideal_i \times g_i) - \left(\left(\sum_{i=1}^{n} ideal_i \times \sum_{i=1}^{n} g_i\right)/n\right)}{\sqrt{\left(\sum_{i=1}^{n} ideal_i^2 - \left(\left(\sum_{i=1}^{n} ideal_i\right)^2/n\right)\right)\left(\sum_{i=1}^{n} g_i^2 - \left(\left(\sum_{i=1}^{n} g_i\right)^2/n\right)\right)}} \quad (5)$$

All the conditional probabilities of NBs are estimated from the training set. A_i is the ith state of a feature A, and count(A_i) is the number of samples whose state is A_i. The conditional probability $P(A_i)$ can be estimated with below equation

$$P(A_i) = \frac{\text{count}(A_i)}{n_T} \quad (6)$$

If A has a parent node B, $P(A_i|B_j)$ can be estimated by

$$P(A_i|B_j) = \frac{\text{count}(A_i, B_j)}{\text{count}(B_j)} \quad (7)$$

The probability of each class is calculated by inference using n features as evidence as follows:

$$P(C|F_1, \ldots, F_n) \quad (8)$$

over a class C and $F_1 \sim F_n$. Using the Bayes theorem [16], we can rewrite it as

$$P(C|F_1, \ldots, F_n) = \frac{P(C)P(F_1, \ldots, F_n|C)}{P(F_1, \ldots, F_n)} \quad (9)$$

In practice, we are only interested in the numerator of the fraction, since the denominator does not affect C. Feature F_i is conditionally independent from the other feature F_j for $j \neq i$, so the probability of a class is described by

$$\begin{aligned} P(C)P(F_1, \ldots, F_n|C) &= P(C)P(F_1|C)P(F_2|C)\ldots P(F_n|C) \\ &= P(C)\prod_{i=1}^{n} P(F_i|C) \end{aligned} \quad (10)$$

```
Prob[m] = { p1, p2, ..., pm}                // prob[] is calculated by NBs
order[m] = {0, 1, 2, ..., m-1}
o-svm[m] = { ma1, ma2, ..., mam}            // o-svm[] is obtained by OVR SVMs
a[m] = {a1, a2, ..., am}                    // ai is a threshold for each OVR SVM

// determine the order of OVR SVMs to evaluate
for (i=0;  i<m;  i++)
  for (j=i+1;  j<m;  j++)
    if (prob[i] < prob[j])  {
      int iTemp = prob[i];  prob[i] = prob[j];  prob[j] = iTemp;
      iTemp = order[i];     order[i] = order[j];  order[j] = iTemp; }

// classify with OVR SVMs according to the subsumption architecture
for (i=0;  i<m;  i++)
  if (o-svm[order[i]] >= a[order[i]])
    return order[i];
return order[0];
```

Fig. 3. Pseudo code for probabilistically ordering OVR SVMs.

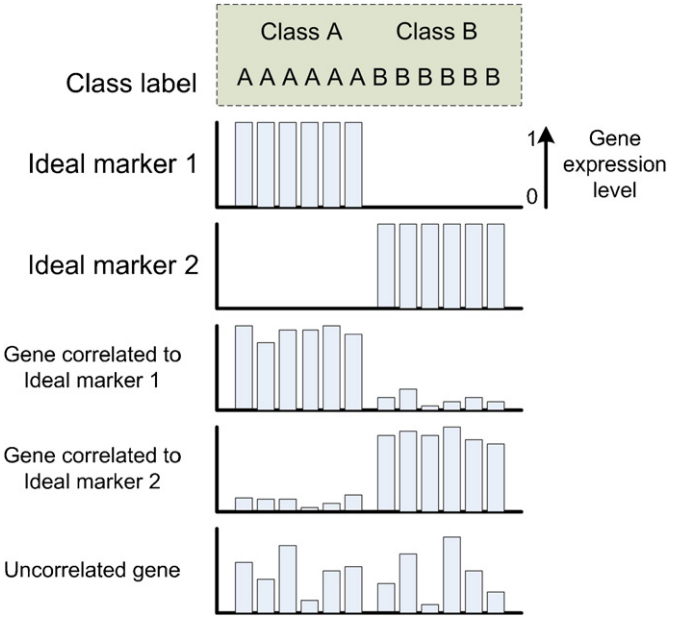


Fig. 4. Negatively correlated features.

4. Experiments

4.1. Experimental results on the GCM cancer dataset

At first, we have verified the proposed method with the GCM cancer dataset consisting of 144 training samples and 54 test samples with 16,063 gene expression levels [21], which is a popular multi-class micro-array dataset. There are 14 different tumor categories including breast adenocarcinoma, prostate, lung adenocarcinoma, colorectal adenocarcinoma, lymphoma, bladder, melanoma, uterine adenocarcinoma, leukemia, renal cell carcinoma, pancreatic adenocarcinoma, ovarian adenocarcinoma, pleural mesothelioma, and central nervous system. Since the dataset provides only a few samples but lots of features, it is a challenging task for many machine learning researchers to construct a competitive classifier.

Ramaswamy et al. [21] and Yeang et al. [25] produced an accuracy of 78% by using OVR SVMs, while Li et al. [15] yielded an

Table 2
Classification accuracy for the GCM cancer dataset

Method (feature #)	Winner-takes-all		Product (SVMs+NBs)	Sum (SVMs+NBs)	ECC (SVMs)	DT (SVMs)	Proposed method
	SVMs (16,063)	NBs (140)					
Accuracy (%)	76.9	74.8	66.7	79.6	77.8	72.2	81.5

accuracy of 63.3%. Statnikov et al. [23] obtained an accuracy of 76.6% for an extended GCM cancer dataset that includes 308 samples of 26 categories. Most of them divided the data as 144 training samples and 54 test samples like the initial setting of Ramaswamy et al., where we also follow it. 140 genes are selected for learning NBs based on Pearson correlation. The linear kernel is used as a basis kernel function of SVMs. All the features of samples are normalized from 0 to 1.

Table 2 shows superior results of the proposed method to several traditional approaches including several sophisticated combining schemes such as winner-takes-all ECC and decision tree (DT). SVMs with the winner-takes-all strategy produced 76.9% classification accuracy, while NBs with Pearson-correlated features yielded an accuracy of 74.8%, individually. The product-based fusion of SVMs and NBs obtained an accuracy of 66.7%, while the sum-based fusion of SVMs and NBs achieved an accuracy of 79.6%. ECC with SVMs yielded an accuracy of 77.8% and DT produced 72.2% classification accuracy. On the other hand, the proposed method produced a classification accuracy of 81.5%, higher than the others. A confusion matrix for the test set is presented in Table 3. From this table, we can see that 100% accuracy has been obtained for lung, colorectal, uterus, and mesothelioma and 83% for CNS, prostate, lymphoma, and leukemia, respectively.

We have categorized samples classified by SVMs and NBs as shown in Table 4. Only two samples were misclassified by both methods, while the proposed method correctly classified four samples failed by SVMs and four samples failed by NBs. One and two ties are occurred by NBs and SVMs, respectively, while all of them are correctly classified by the proposed method. Table 5 shows some examples that can be misclassified when using an individual classifier, but not by the proposed method. This result means that the proposed method is useful to improve the classification performance by effectively combining two different classification methods. Table 6 summarizes some related works that used the GCM cancer dataset in their experiments, where the proposed method ranks in high among them.

Table 3
Confusion matrix

%	0	1	2	3	4	5	6	7	8	9	10	11	12	13	*n*
0	75										25				4
1		83										17			6
2			100												4
3				100											4
4			17		83										6
5			33			67									3
6							50				50				2
7								100							2
8		17							83						6
9								33		67					3
10				33		33					34				3
11				25			25					50			4
12													100		3
13														100	4
n	3	6	6	6	5	3	2	3	5	2	3	3	3	4	54

0: breast, 1: prostate, 2: lung, 3: colorectal, 4: lymphoma, 5: bladder, 6: melanoma, 7: uterus_adeno, 8: leukemia, 9: renal, 10: pancreas, 11: ovary, 12: mesothelioma, 13:CNS.

Table 4
Analysis of classification results

(Correct/incorrect): proposed method	SVMs		
	O	X	Tie
NBs			
O	32 (32/0)	6 (4/2)	2 (2/0)
X	7 (4/3)	2 (0/2)	0
Tie	1 (1/0)	4 (1/3)	0

4.2. Comparisons on other benchmark datasets

In order to verify the applicability of the proposed method, we have conducted additional experiments on four other popular cancer datasets that include two brain datasets and two leukemia datasets as shown in Table 7. They had 3–5 distinct diagnostic categories, 50–90 samples, and 5327–11,225 gene expression levels. Leave-one-out cross-validation, which is a popular choice for the split ratio of the dataset into a training set and a test set, were executed with the datasets. After dividing datasets, a set of informative genes are selected according to the Pearson correlation using the training set to train NBs.

Table 8 shows competitive classification results for the benchmark datasets. Since the product scheme is sensitive to noise, it shows lower performance than others, even individual classifiers. In most cases, the proposed method yields higher classification accuracy than the others, especially including winner-takes-all, ECC and DT.

Table 5
Test examples

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	Class label
3rd															
SVM	0.3	−1.1	−1.3	0.3	−1.4	−1.6	−1.2	−0.1	−1.5	−1.3	−1.7	−2.1	−2.1	−1.4	0
NB	0.89	0	0	0.04	0	0.07	0	0	0	0.01	0.01	0.02	0	0	
16th															
SVM	0.3	−1.1	−1.3	0.3	−1.4	−1.6	−1.2	−0.1	−1.5	−1.3	−1.7	−2.1	−2.1	−1.4	3
NB	0.01	0	0	0.71	0	0.01	0	0	0	0.01	0	0	0	0	
27th															
SVM	−1.2	−0.7	−1.2	−0.8	−1.3	−0.1	−1.0	−1.5	−1.1	−0.9	−0.7	−1.0	−1.7	−1.0	5
NB	0.01	0	0	0.01	0	0.01	0	0	0	0	0	0	0	0	
39th															
SVM	−1.3	−1.6	−1.3	−1.0	−1.7	−1.1	−1.2	−0.0	−1.2	0.1	−2.4	−0.9	−1.5	−1.0	9
NB	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	

Table 6
Related works on the GCM cancer dataset

Author	Data set	# Genes	# Classes	# Samples	Method	Accuracy (%)
Ramaswamy et al. [21]	GCM	16,063	14	198 (144/54)	OVR SVM	77.8
Yeang et al. [25]	GCM	16,063	14	190 (144/46)	OVR SVM	78.2
Ooi and Tan [19]	GCM	16,063	14	198 (144/54)	GA/MLHD	82.0
	NCI60	1000	9	61 (LOOCV)	GA/MLHD	85.4
Li et al. [15]	GCM	16,063	14	198 (144/54)	SVM random	63.3
	NCI60	1123	9	60 (LOOCV)	SVM random	66.7
Tan et al. [24]	DLBCL	4026	6	88 (58/30)	HC-k-TSP	83.3
	GCM	16,063	14	190 (144/46)	HC-k-TSP	67.47
Deutsch [7]	GCM	16,063	14	190 (144/46)	GA/SVM	77.9
Statnikov et al. [23]	GCM	15,009	26	308 (10-fold CV)	MC-SVM	76.6

Table 7
Other benchmark cancer datasets

Dataset name	# Genes	# Classes	# Samples	Author
Brain_Cancer_1 (B1)	5920	5	90	Pomeroy et al. [20]
Brain_Cancer_2 (B2)	10,367	4	50	Nutt et al. [18]
Leukemia_Cancer_1 (L1)	5327	3	72	Golub et al. [9]
Leukemia_Cancer_2 (L2)	11,225	3	72	Armstrong et al. [2]

Table 8
Classification results for the datasets

Dataset	OVR SVMs	NBs	Product	Sum	ECC	DT	Proposed method
B1	78 (87%)	69 (77%)	64 (71%)	68 (76%)	78(87%)	79(88%)	80 (89%)
B2	35 (70%)	32 (64%)	30 (60%)	35 (70%)	35(70%)	35(70%)	38 (76%)
L1	67 (93%)	68 (94%)	67 (93%)	68 (94%)	67(93%)	67(93%)	68 (94%)
L2	70 (97%)	62 (86%)	62 (86%)	64 (89%)	70(97%)	70(97%)	71 (99%)

5. Concluding remarks

Multi-class classification is a challenging task in pattern recognition, where various approaches have been investigated especially using SVMs. Since SVM is a basically binary classifier, it is necessary to formulate decomposition and reconstruction methods. In this paper, we have proposed a hybrid classifier that integrates SVMs and NBs learned based on the OVR scheme. Several popular multi-class benchmark datasets in bioinformatics including the GCM cancer dataset are used to verify the proposed method. Since the dataset consists of huge number of genes, we have reduced the dimensionality by using a feature selection method with Pearson correlation. The proposed method has showed better performance than OVR SVMs and NBs when working individually or combined by product and sum strategies. As the future work, we will demonstrate the proposed method with other popular benchmark datasets of multi-class, and study about the thresholds for better performance.

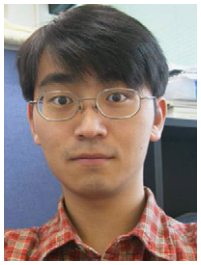
Acknowledgment

This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Advancement) IITA-2008-(C1090-0801-0011).

References

[1] C. Angulo, et al., K-SVCR. A support vector machine for multi-class classification, Neurocomputing 55 (1–2) (2003) 57–77.
[2] S. Armstrong, et al., MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia, Nat. Genet. 30 (1) (2002) 41–47.
[3] E. Bredensteiner, K. Bennett, Multicategory classification by support vector machines, Comput. Optim. Appl. 12 (1) (1999) 53–79.
[4] R. Brooks, A robust layered control system for a mobile robot, IEEE J. Robotics Autom. 2 (1) (1986) 14–23.
[5] S.-B. Cho, J. Ryu, Classifying gene expression data of cancer using classifier ensemble with mutually exclusive features, Proc. IEEE 90 (11) (2002) 1744–1753.
[6] K. Crammer, Y. Singer, On the learnability and design of output codes for multiclass problems, Mach. Learn. 47 (2–3) (2002) 201–233.
[7] J. Deutsch, Evolutionary algorithms for finding optimal gene sets in microarray prediction, Bioinformatics 19 (1) (2003) 45–52.
[8] T. Gestel, et al., Multiclass LS-SVMs: moderated outputs and coding-decoding schemes, Neural Process. Lett. 15 (1) (2002) 45–58.
[9] T. Golub, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (15) (1999) 531–537.
[10] C. Hsu, C. Lin, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Networks 13 (2) (2002) 415–425.
[11] I. Inza, et al., Feature subset selection by genetic algorithm and estimation of distribution algorithms. A case study in the survival of cirrhotic patients treated with TIPS, Artif. Intell. Med. 23 (2) (2001) 187–205.
[12] J.-Y. Koo, et al., Structured polychotomous machine diagnosis of multiple cancer types using gene expression, Bioinformatics 22 (8) (2006) 950–958.
[13] L. Kuncheva, A theoretical study on six classifier fusion strategies, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2) (2002) 281–286.
[14] Y. Lee, C. Lee, Classification of multiple cancer types by multicategory support vector machines using gene expression data, Bioinformatics 19 (9) (2003) 1132–1139.
[15] T. Li, et al., A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression, Bioinformatics 20 (15) (2004) 2429–2437.
[16] J. Liu, et al., An improved naïve Bayesian classifier technique coupled with a novel input solution method, IEEE Trans. Syst. Man Cybern.—Part C: Appl. Rev. 31 (2) (2001) 249–256.
[17] J. Liu, et al., Multiclass cancer classification and biomarker discovery using GA-based algorithms, Bioinformatics 21 (11) (2005) 2691–2697.
[18] C. Nutt, et al., Gene expression-based classification of malignant gliomas correlates better with survival than histological classification, Cancer Res. 63 (7) (2003) 1602–1607.
[19] C. Ooi, P. Tan, Genetic algorithms applied to multi-class prediction for the analysis of gene expression data, Bioinformatics 19 (1) (2003) 37–44.
[20] S. Pomeroy, et al., Prediction of central nervous system embryonal tumour outcome based on gene expression, Nature 415 (24) (2002) 437–442.
[21] S. Ramaswamy, et al., Multiclass cancer diagnosis using tumor gene expression signatures, Proc. Nat. Acad. Sci. 98 (26) (2001) 15149–15154.

[22] D. Sebald, J. Bucklew, Support vector machines and the multiple hypothesis test problem, IEEE Trans. Signal Process. 49 (11) (2001) 2865–2872.
[23] A. Statnikov, et al., A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis, Bioinformatics 21 (5) (2005) 631–643.
[24] A. Tan, et al., Simple decision rules for classifying human cancers from gene expression profiles, Bioinformatics 21 (20) (2005) 3896–3904.
[25] C.-H. Yeang, et al., Molecular classification of multiple tumor types, Bioinformatics 17 (Suppl. 1) (2001) 316–322.



Jin-Hyuk Hong received the BS and MS degrees in computer science from Yonsei University, Seoul, Korea, in 2002 and 2004, respectively. Since 2004, he is a Ph.D. student in the Department of Computer Science, Yonsei University. His research interests include evolutionary computation, conversational intelligent agent and game strategy generation.

Sung Bae Cho received the BS degree in computer science from Yonsei University, Seoul, Korea, in 1988 and the MS and Ph.D. degrees in computer science from KAIST (Korea Advanced Institute of Science and Technology), Taejeon, Korea, in 1990 and 1993, respectively. He has worked as a Member of the Research at the Center for Artificial Intelligence Research at KAIST from 1991 to 1993. He was an Invited Researcher of Human Information Processing Research Laboratories at ATR (Advanced Telecommunications Research) Institute, Kyoto, Japan from 1993 to 1995, and a visiting scholar at University of New South Wales, Canberra, Australia in 1998. Since 1995, he has been an associate professor in the Department of Computer Science, Yonsei University. His research interests include neural networks, pattern recognition, intelligent man–machine interfaces, evolutionary computation, and artificial life. Dr. Cho was awarded outstanding paper prizes from the IEEE Korea Section in 1989 and 1992, and another one from the Korea Information Science Society in 1990. He was also the recipient of the Richard E. Merwin prize from the IEEE Computer Society in 1993. He was listed in Who's Who in Pattern Recognition from the International Association for Pattern Recognition in 1994, and received the best paper awards at International Conference on Soft Computing in 1996 and 1998. Also, he received the best paper award at World Automation Congress in 1998, and listed in Marquis Who's Who in Science and Engineering in 2000 and in Marquis Who's Who in the World in 2001. He is a Member of the Korea Information Science Society, INNS, the IEEE Computer Society, and the IEEE Systems, Man, and Cybernetics Society.