A Biclustering Based Classification Framework for Cancer Diagnosis and Prognosis

Baljeet Malhotra and Guohui Lin*

Department of Computing Science,
University of Alberta,
Edmonton, Alberta, Canada T6G 2E8
Emails: {baljeet,ghlin}@cs.ualberta.ca

Abstract. In gene expression microarray data analysis, biclustering has been demonstrated to be one of the most effective methods for discovering gene expression patterns under various conditions. We present in this study a framework to take advantage of the homogeneously expressed genes in biclusters to construct a classifier for sample class membership prediction. Extensive experiments on 8 real cancer microarray datasets (4 diagnostic and 4 prognostic) show that our proposed classifier performed superior in both cancer diagnosis and prognosis, the latter of which was regarded quite difficult previously. Additionally, our results demonstrate that sample classification accuracy can serve as a good subjective quality measure for biclusters.

1 Introduction

The advance of high-throughput hybridization microarray technology provides the opportunity to measure the expression levels of thousands of genes simultaneously, thus to present a snapshot of the transcription levels within the cell. Such a technology enables researchers to look at the cellular systems globally, for example, to improve our understanding on the disease related processes, yet also challenges us on effectively analyzing the vast volume of measured data such that key features of the cellular systems can be uncovered.

One of the major current applications of gene expression microarrays, particularly the high-density oligonucleotide arrays such as the Affymetrix GeneChip oligonucleotide (Affy) arrays, is cancer diagnosis and prognosis. The underlying principle for this application (and many other applications) is that, two cells with dramatically different biological characteristics, such as a normal cell versus a cancerous cell from the same tissue, are expected to have different gene expression profiles. However, it is important to realize that the majority of the active cellular mRNA is not affected by the differences. In other words, a dramatic biological difference does have a gene expression level manifestation, but the set of genes that is involved can be rather small. The microarray classification is to partition the arrays (also called samples, or chips, or conditions) such that there are an extremely larger-than-expected number of genes *sharply* separating the classes.

* Correspondence author.

These genes that sharply separate the classes are referred to as *informative* or *discriminatory* genes, or *biomarkers*. Since these genes expressed differentially under different conditions, they can be selected to compose gene expression profiles for the purpose of class prediction, upon the arrival of a new sample. However, some early works on classification and/or class discovery are rather direct in that their focus is on the sample partition (not prediction).[1] Some others investigate two-way clustering of both genes and samples for defining sample classes and the class associated gene identification.[2] Note that sample partitioning requires homogeneous expression for all the genes while gene clustering assumes homogeneous expression of genes across all samples. With the increased understanding that not all genes express similarly in all samples, an alternate clustering framework, which produces local models, has been proposed to group genes and samples simultaneously, the so-called *biclustering*, which is also known (in several other areas of studies) as co-clustering, bi-dimensional clustering, and subspace clustering.[3]

In the literature, there are several types of biclusters been defined and investigated [4–9]. Among them, *constant*,[10] *additive*,[6] and *multiplicative*[9, 5] are three most studied types. Associated with different types of biclusters, various algorithms for finding them have been proposed,[10, 6, 9] together with some theoretical studies on computational complexity.[3, 6] In these works, several bicluster quality measures have been examined, using methods such as value of the merit function defined for biclusters, statistical significance of the solution measured against the null hypothesis, and comparison against known solutions.[3, 6] Note that the first two methods emphasize the numerical quality of the identified biclusters, while the third can incorporate existing biological knowledge such as gene functional annotation, gene co-regulation, and sample class membership.[3] For example, several works reviewed by Madeira and Oliveira[3] examine the relation between biclusters and sample class memberships.[5, 7] Unfortunately, it is unclear from the context on how biclusters can be used for sample class membership *prediction*, or sample classification, which is our main target in this study.

In this paper, we present a detailed framework for sample classification using biclusters, and we design experiments to show that good quality biclusters can be taken advantage for human cancer diagnosis and prognosis. Furthermore, the experimental results demonstrate that sample classification accuracy can serve as a good quality measurement for the discovered biclusters, disregarding their types.

The rest of paper is organized as follows: In the next section, we briefly introduce the concept of biclusters, particularly the constant and the additive, and two existing algorithms for finding them. With the discovered biclusters, we present the framework on using the genes in the biclusters for sample classification within the leave-one-out cross validation (LOOCV) scheme. Section 3 presents the cancer (diagnosis and prognosis) microarray datasets included in this study, and our experimental results on them. Section 4 contains our discussion on both the classification framework and computational results. We conclude the paper in Section 5.

2 Methods

We use A to denote the gene expression data matrix in the study. In this case, A is an $n \times m$ matrix, with n being the number of genes and m being the number of samples.

The entry a_{ij} records the expression level of the i-th gene in the j-th sample. Note that the order of genes and the order of samples are (normally) arbitrary and irrelevant in this study. Given a subset of genes $I \subseteq \{1, 2, \ldots, n\}$ and a subset of samples $J \subseteq \{1, 2, \ldots, m\}$, A_{IJ} denotes the sub-matrix of A by removing genes not in I and samples not in J.

Different from clustering (on genes or samples) which seeks for homogeneous gene expression (across all samples, or for all genes, respectively), biclustering performs clustering in the two dimensions simultaneously, and thus to produce local models contrast to global models produced by clustering. A bicluster is defined by a pair of a gene subset I and a sample subset J, expecting that genes in I have similar behavior across the samples in J. The notion of "similar behavior" can be characterized in several ways.[3] In this paper we are particularly interested in two types of biclusters: constant and additive.

A *perfect* constant bicluster is a sub-matrix A_{IJ} in which all entries are equal, that is, $a_{ij} = \mu$, for all $i \in I$ and $j \in J$. A *perfect* additive bicluster is a sub-matrix A_{IJ} with *coherent* values, which can be expressed as $a_{ij} = \mu + \alpha_i + \beta_j$, where α_i is the *adjustment* for the i-th gene and β_j is the *adjustment* for the j-th sample. Clearly, a perfect constant bicluster is a special case of a perfect additive bicluster. Although these "ideal" biclusters can be found in some expression matrices, in real data, they are masked by noise.

Many algorithms have been proposed for discovering these two types of bi-clusters or alike. In this paper, we employ two of them, to be detailed next. In their seminal work,[10] Cheng and Church defined a bicluster to have a high *mean squared residue score*, which is used as a measure of the coherence of the genes and samples in the bicluster. Given a bicluster A_{IJ}, set $a_{iJ} = \frac{1}{|J|}\sum_{j\in J} a_{ij}$, $a_{Ij} = \frac{1}{|I|}\sum_{i\in I} a_{ij}$, and $a_{IJ} = \frac{1}{|I||J|}\sum_{i\in i, j\in J} a_{ij}$. The *residue score* of entry a_{ij} in the bicluster is $a_{ij} - a_{iJ} - a_{Ij} + a_{IJ}$. The *mean squared residue score* of A_{IJ}, denoted as $H(I, J)$, is calculated as

$$H(I, J) = \sum_{i\in I, j\in J} (a_{ij} - a_{iJ} - a_{Ij} + a_{IJ})^2.$$

A_{IJ} is a δ-bicluster if $H(I, J) \leq \delta$ for some $\delta \geq 0$.

Clearly, a 0-bicluster is a perfect constant bicluster. Cheng and Church proposed several heuristic algorithms to discover δ-biclusters by removing rows and columns from the original matrix.[10] It is worth mentioning that their proposed algorithms have a tendency to find constant biclusters, but not necessarily other types of biclusters such as additive or multiplicative. The particular algorithm we employ in this study is the *Multiple Node Deletion* (MND) algorithm, which iteratively removes genes whose contributing residue scores (defined as $\frac{1}{|J|}\sum_{j\in J} (a_{ij} - a_{iJ} - a_{Ij} + a_{IJ})^2$) are greater than $\alpha H(I, J)$, or when no such genes, only the gene with the highest such score, and samples whose contributing residue scores (defined as $\frac{1}{|I|}\sum_{i\in I} (a_{ij} - a_{iJ} - a_{Ij} + a_{IJ})^2$) are greater than $\alpha H(I, J)$, or when no such samples, only the sample with the highest such score, until $H(I, J)$ does not exceed δ. We denote this algorithm as MND(δ, α) for the particular pair of δ and α.

Recently, Liu and Wang proposed several algorithms for finding multiple (may be overlapping) *maximum similarity* biclusters,[6] which include constant and addi-

tive ones. Given I and J, and a reference gene $i^* \in I$, finding a maximum similarity bicluster within I and J is to find a subset of genes $I' \subseteq I$ and a subset of samples $J' \subseteq J$ such that the *distances* between the reference gene i^* and genes in I' are minimized. In more details, define $d_{ij} = |a_{ij} - a_{i^*j}|$, and we want to discard those large d_{ij}'s to achieve the target bicluster. To this purpose, define the average difference as $\bar{d} = \frac{1}{|I||J|} \sum_{i \in I, j \in J} d_{ij}$. With a threshold α, a similarity matrix S_{IJ} is defined for A_{IJ} in which $s_{ij} = 0$ if $d_{ij} \geq \alpha\bar{d}$, or otherwise $s_{ij} = 1 - d_{ij}/\alpha\bar{d} + \beta$ (where $\beta \geq 0$ is a *bonus* for small d_{ij}'s). Define the similarity score of the i-th gene in S_{IJ} as $s_{iJ} = \sum_{j \in J} s_{ij}$, the similarity score of the j-th sample in S_{IJ} as $s_{Ij} = \sum_{i \in I} s_{ij}$, and the similarity score of matrix A_{IJ} as $s_{IJ} = \min\{\min_{i \in I} s_{iJ}, \min_{j \in J} s_{Ij}\}$. The particular algorithm we employ in this study is the MSB algorithm, which starts with the whole matrix A, repeatedly deletes the gene or the sample whose similarity score is the currently smallest to obtain $n + m - 1$ biclusters, and returns the one having the maximum similarity score s_{IJ} while its average similarity score $\bar{s}_{IJ} = \frac{1}{|I||J|} \sum_{i \in I, j \in J} s_{ij}$ is at least γ. We denote this algorithm as MSB(γ, α, β) for the three associated parameters.

2.1 Bicluster Generation

We use two algorithms, MND(δ, α) and MSB(γ, α, β), to generate biclusters in the (*training*) dataset, in which every sample has a known class membership. For each algorithm, a range is pre-determined for every parameter, resulting in a number of distinct settings. First of all, we partition the expression matrix A (the training dataset) into submatrices which have the same set of genes but each contains only the samples from a particular class. (The number of such sub-matrices is equal to the number of classes in the dataset.) Separately or together, every setting of the biclustering algorithms is run on these sub-matrices to generate one bicluster per class.

In MND(δ, α) algorithm, δ was chosen based on a trial and error policy. With this policy, first, we ran MND(δ, α) algorithm (with some random δ value) on the whole dataset while noting down the residue score ($H(I, J)$) and the size of the generated bicluster. When the size of the gene set in the generated bicluster was less than half of the total number of genes, we chose that particular δ value as the initial value. Afterwords, we ran the algorithm multiple times while incrementally decreasing the δ value. For example, if the initial δ value for a given dataset was 700, we subsequently ran MND(δ, α) algorithm with δ values being 600, 500, and 400, and so on. Under this policy of parameter setting, the size of the gene set in the generated biclusters varied from 30% to 10% of the total number of genes in the whole dataset. Parameter α was kept at 1.1 throughout the experiments.

For MSB(γ, α, β) algorithm, parameters α, β, and γ were chosen based on the original recommendations[6], where the authors suggested that $\alpha \in [0.2, 0.4]$, $\beta \in [0.0, 0.5]$, and $\gamma \in [\beta + 0.7, \beta + 0.9]$. We tried $\alpha = 0.3, 0.4$, and fixed β at 0.4 and γ at 1.2. With these settings the sizes of the gene sets of the generated biclusters varied from 20% to 8% of the total number of genes in the whole dataset. For each setting of (γ, α, β), (a maximum of) 5 reference genes were randomly selected from the gene pool.

2.2 Distance Calculation

The generated biclusters are considered as important and the genes in them are believed to strongly correlate to the sample classes. We take all the genes included in these top quality biclusters for calculating distances between a testing sample and the sample classes in the training dataset. Assume these genes form a set I and the training sample set is J. For each sample $j \in J$, the distance between testing sample s and sample j is calculated as the normalized L_1 distance using gene set I:

$$d_{L_1}(s, j) = \frac{1}{|I|} \sum_{i \in I} |a_{is} - a_{ij}|.$$

Note that a sample not in J has no distance to testing sample s. The distance between testing sample s and a sample class is defined as the *average* distance over all the samples in the class which have distances to s. The above L_1 distance can be substituted by other distance measures such as the euclidean distance.

2.3 Classification and LOOCV Accuracy

Given all the discovered biclusters which define the gene set used in the distance calculation, whenever a testing sample arrives, we can calculate its distance to every sample class in the (training) dataset. The label of the closest class to the testing sample is taken as the predicted class label for the testing sample. In our experiment, we adopt the leave-one-out cross validation (LOOCV) scheme to calculate the classification accuracy. At each iteration, one sample is selected as the testing sample whose class membership is blinded to the classifier. Using the rest of the samples, biclustering algorithms are run to generate the target biclusters and the subsequent genes used in the distance calculation. One correct prediction is arrived when the predicted class label is the same as the true one. The LOOCV scheme iterates through all samples and the percentage of correct predictions is the LOOCV classification accuracy.

3 Experimental Results

3.1 Overview

All experiments were conducted in Matlab environment. We have coded both algorithms, MND(δ, α) and MSB(γ, α, β), ourselves and thoroughly tested their correctness. For example, using the same datasets in their original paper, our coded algorithms were tested on to generate biclusters, which were compared to the biclusters generated by the original authors. A test case is considered successful only if these two sets of biclusters matched with each other. The correctness is guaranteed by 100% matching results in several test cases.

Afterwords, complete LOOCV sample classification was performed, using these two algorithms either separately or jointly, on several real cancer gene expression microarray datasets, for either diagnosis or prognosis purpose. The classification accuracies were reported and compared to the previously achieved best accuracies on the individual datasets.

3.2 Cancer Gene Expression Datasets

We have used 4 cancer diagnosis datasets and 4 prognosis datasets in our experiments, listed as follows.

Diagnostic Datasets AML-ALL Leukemia dataset[11] consists of 72 samples in two classes: *acute lymphoblastic leukemia* (ALL) and *acute myeloid leukemia* (AML). The gene expression levels were measured using Affymetrix high-density oligonucleotide arrays containing $7,129$ probes (from $6,817$ human genes), 47 samples of ALL (38 B-cell ALL and 9 T-cell ALL) and 25 samples of AML. After filtering the dataset contains $3,571$ genes. In one of our previous works[12], we achieved an LOOCV classification accuracy 98.60%, which is the best known on this dataset.

Lung Cancer dataset[13] is used for classification between *malignant pleural mesothelioma* (MPM) and *adenocarcinoma* (ADCA) of the lung. There are 181 tissue samples (31 MPM and 150 ADCA), on 12,533 genes. We do not have any known LOOCV result on this dataset.

The Brain tumor dataset consists of 50 high-grade glioma samples of which 28 are *glioblastomas* and 22 are *anaplastic oligodendrogliomas*.[14] Glioblastomas and anaplastic oligodendrogliomas samples are further classified into *classic* and *non-classic* tumors (14 and 14, 7 and 15, respectively). This dataset contains 12,625 genes. The best ever achieved LOOCV classification accuracy on this dataset is 80%.[15]

The Carcinomas dataset (U95a GeneChip) contains 174 samples in 11 classes: *prostate*, *bladder*/*ureter*, *breast*, *colorectal*, *gastroesophagus*, *kidney*, *liver*, *ovary*, *pancreas*, *lung adenocarcinomas*, and *lung squamous cell carcinoma*, which have 26, 8, 26, 23, 12, 11, 7, 27, 6, 14, and 14 samples, respectively.[16] Each sample originally contained 12,533 genes. We preprocessed the dataset as described in Su *et al.*[16] to include only those probe sets whose maximum hybridization intensity is ≥ 200 in at least one sample; Subsequently, all hybridization intensity values ≤ 20 were raised to 20, and the values were log transformed. After preprocessing, we obtained a dataset of 9,183 genes. The best ever achieved LOOCV classification accuracy on this dataset is 93.6%.[15]

Prognostic Datasets Breast Cancer (training) dataset[17] contains 78 patient samples, 34 of which are from patients who had developed distance metastases within 5 years (labeled as *relapse*), the rest 44 samples are from patients who remained healthy from the disease after their initial diagnosis for interval of at least 5 years (labeled as *non-relapse*). The original dataset contains 24,481 genes. Our version of dataset contains only 23,625 genes and 32 relapse samples and 44 non-relapse samples. The authors applied a selection scheme on genes and constructed a classifier based on the correlation coefficient to good prognosis templates and poor prognosis templates. The achieved LOOCV classification accuracy on this dataset is 73%.

AML-Leukemia is a subset of the above described AML-ALL Leukemia dataset[11], which contains $7,129$ probes (from $6,817$ human genes) and 15 samples of AML. 8 treatments failed and the other 7 were successful. There is no LOOCV result for this dataset.

Central Nervous System dataset[18] is used to analyze the outcome of the treatment. Survivors are patients who are alive after treatment whiles the failures are those who succumbed to their disease. The dataset contains 60 patient samples, 21 are *survivors* and 39 are *failures*, on 7,129 genes. The authors selected a subset of genes to construct a KNN-classifier and achieved a LOOCV classification accuracy of 78%.

Prostate Cancer dataset[19] for prediction of clinical outcome contains 21 patients were evaluable with respect to recurrence following surgery with 8 patients having *relapsed* and 13 patients having remained relapse free (*non-relapse*) for at least 4 years. The dataset contains 12,600 genes. The authors selected a subset of genes to construct a KNN-classifier and achieved a LOOCV classification accuracy of 90%.

3.3 LOOCV Classification Accuracies

Table1 summarizes our results. The size column records the size of an individual dataset using the number of genes and the numbers of samples in all the classes. On each dataset, the previously best classification accuracy, to the best of our knowledge, is added for comparison purpose. We have three LOOCV classification accuracies, by only MND(δ, α) algorithm, by only MSB(γ, α, β), and by both of them jointly.

Table 1. The LOOCV classification accuracies achieved by our bicluster-based methods, compared with the previously achieved best accuracies, on the eight cancer gene expression microarray datasets. The bold ones are the currently best LOOCV classification accuracies.

Dataset		Prev. Best	Our Accuracies (%)		
Name	Size	Accuracy (%)	MND	MSB	MND+MSB
ALL-AML Leukemia	3,371 $\times$ {47, 25}	[12]98.60	97.22	**98.66**	97.22
Lung Cancer	12,533 $\times$ {150, 31}	-	**88.95**	84.53	88.95
Brain Tumor	12,625 $\times$ {14, 14, 7, 15}	[15]80.00	88.00	**92.00**	88.00
Carcinomas	9,183 $\times$ {26, 8, 26, . . .}	[15]93.60	91.95	**96.55**	91.95
Breast Cancer	23,625 $\times$ {32, 44}	[17]73.00	53.94	71.05	53.94
Leukemia-AML	7,129 $\times$ {8, 7}	-	**100.00**	100.00	100.00
Central Nervous System	7,129 $\times$ {39, 21}	[18]**78.00**	40.00	53.33	40.00
Prostate Cancer	12,600 $\times$ {8, 13}	[19]90.00	80.95	**95.23**	80.95

On six of the eight dataset for which we know the previously best LOOCV results, 4 of them are updated by our proposed method (in bold in Table 1). In particular, on the Carcinomas dataset, the detailed prediction results by MSB(0.3, 0.45, 1.2), using three randomly chosen reference genes, on all the classes are recorded in Table 2, where the correct predictions are in bold.

On the Breast Cancer dataset, our method performed competitively, 71.05% versus 73.00%. On the last Central Nervous System dataset, our method did not perform satisfactorily. It is worth noting that the small Leukemia-AML prognostic dataset was considered challenging for computational prognosis previously[11]. Our method achieved the perfect result on this small dataset.

Table 2. The detailed prediction results by MSB(0.3, 0.45, 1.2), using three randomly chosen reference genes, on all the classes in the Carcinomas dataset, where the correct predictions are in bold.

# Samples		P	BU	B	C	G	K	LI	O	PA	LA	LS
Prostate (P)	26	**26**										
Bladder/Ureter (BU)	8		**8**									
Breast (B)	26	1	1	**23**	1							
Colorectal (C)	23				**23**							
Gastroesophagus (G)	12					**11**					1	
Kidney (K)	11						**11**					
Liver (LI)	7							**7**				
Ovary (O)	27			1					**26**			
Pancreas (PA)	6									**6**		
Lung Adeno. (LA)	14										**14**	
Lung Squamous (LS)	14				1							**13**

4 Discussion

4.1 Gene Selection

In the past several years, many sample classification algorithms have been proposed, most of which deal with the dimensionality issue (that is, tens of thousands of genes versus only tens of samples) through a step called gene selection. Essentially, various mechanisms have been set up to identify the most discriminatory genes, which express substantially different under different conditions, followed by classifier construction based on the selected genes.

Our classification method based on discovered biclusters may also be regarded as one of the kind, in that our selected genes are those that are included in the discovered biclusters. Nevertheless, our "gene selection" is very different from the existing ones in principle. Within the biclustering context, we partition the whole expression matrix into sub-matrices such that each contains only those samples in one sample class. The employed biclustering algorithms uncovered those genes that strongly correlate to the class. Therefore, using them in distance calculation is adequate and when the testing sample does belong to the particular class, the distance is expected to be small, or large otherwise.

4.2 Using Class-Dependent Genes Only

We have also tested the distance calculation between the testing sample and a particular class by using only those genes that are included in the biclusters generated for that class. The intention was similar in that when the testing sample belongs to this class, the calculated distance is expected small, or large otherwise. However, the computational results show that such a scheme is inferior, though not much, to the scheme of using all the occurring genes in the distance calculation. We thus chose not to report this set of results.

4.3 The Number of Biclusters

For each class, MND(δ, α) algorithm generated only one bicluster, and MSB(γ, α, β) algorithm generated no more than 5 biclusters. The percentage of genes occurring in these biclusters is roughly 30% to 8% of the total number of genes in the whole dataset. We have also tested to generate many more biclusters, by changing the parameter setting, and then to select a few of them for distance calculation. It turned out that the latter did not perform better, while increased the complexity.

4.4 The Size of Dataset

Most of the running time was consumed by the biclustering algorithms. The problem became more sever with increasing dataset size. With the tens of thousands of genes, the bottleneck is the class size, i.e., the number of samples in the particular class. We experienced some delays on several datasets, such as the diagnostic Lung Cancer dataset and the prognostic Breast Cancer dataset, of which the class sizes are relative large. Note that in the LOOCV scheme, the biclustering algorithms were run for a huge number of times. For example, on the diagnostic Lung Cancer dataset, each algorithm was run for 150 times on a dataset of size 12,533 $\times$ 149 and for 31 times on a dataset of size 12,533 $\times$ 30. When the class sizes are all relatively small, such as the diagnostic Carcinomas dataset (9,183 genes, the maximum class size is 27), the computation was quickly done.

5 Conclusions

In this paper, we presented formally a sample classification framework using the discovered biclusters. The extensive experiments demonstrated that the top ranked constant biclusters generated by two previously proposed algorithms can be taken advantage for the sample classification purpose. As a byproduct, the results demonstrated that sample classification accuracy can serve as an effective and biologically meaningful measurement for the bicluster quality, contrast to previously proposed measures that largely look at the numerical aspects matching to the bicluster definitions. Our proposed sample classification method is a generic framework, in that any biclustering algorithms for finding various types of biclusters can be plugged in.

Some of our future work subjects include investigating which type(s) of biclusters are more helpful for cancer diagnosis and prognosis purposes, better criteria for bicluster selection, better use of the genes included in the selected biclusters, a substantial comparative study to other most advanced classification algorithms, and the limit of our framework in terms of the dataset class number.

Acknowledgments

This research is supported in part by CFI, NSERC.

References

1. A. Ben-Dor, N. Friedman, and Z. Yakhini. Class discovery in gene expression data. In *Proceedings of RECOMB 2001*, pages 31–38, 2001.

2. U. Alon, N. Barkai, D. A. Notterman, and *et al.* Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences USA*, 96:6745–6750, 1999.
3. S. C. Madeira and A. L. Oliveira. Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1:24–45, 2004.
4. H. Cho, I. S. Dhillon, Y. Guan, and S. Sra. Minimum sum-squared residue coclustering of gene expression data. In *Proceedings of the Fourth SIAM International Conference on Data Mining*, 2004.
5. Y. Klugar, R. Basri, J. T. Chang, and M. Gerstein. Spectral biclustering of microarray data: Coclustering genes and conditions. *Genome Research*, 13:703–716, 2003.
6. X. Liu and L. Wang. Computing the maximum similarity bi-clusters of gene expression data. *Bioinformatics*, 23:50–56, 2007.
7. A. Tanay, R. Sharan, and R. Shamir. Discovering statistically significant biclusters in gene expression data. *Bioinformatics*, 18:S136–S144, 2002.
8. J. Yang, W. Wang, H. Wang, and P. Yu. Enhanced biclustering on expression data. In *Proceedings of the Third IEEE Conference on Bioinformatics and Bioengineering*, pages 321–327, 2003.
9. L. Lazzeroni and A. Owen. Plaid models for gene expression data. *Statistica Sinica*, 12:61–86, 2002.
10. Y. Cheng and G. M. Church. Biclustering of expression data. In *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology (ISMB 2000)*, pages 93–103, 2000.
11. T. R. Golub, D. K. Slonim, P. Tamayo, and *et al.* Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, 286:531–537, 1999.
12. Z. Cai, L. Xu, Y. Shi, M. R. Salavatipour, R. Goebel, and G.-H. Lin. Using gene clustering to identify discriminatory genes with higher classification accuracy. In *Proceedings of IEEE The 6th Symposium on Bioinformatics and Bioengineering (IEEE BIBE 2006)*, pages 235–242, Washington D.C., USA, October 16-18, 2006, 2006.
13. G. J. Gordon, R. V. Jensen, L.-L. Hsiao, and *et al.* Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer Research*, 62:4963–4967, 2002.
14. C. L. Nutt, D. R. Mani, R. A. Betensky, and *et al.* Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Research*, 63:1602–1607, 2003.
15. Z. Cai, R. Goebel, M. R. Salavatipour, and G.-H. Lin. Selecting dissimilar genes for multi-class classification, an application in cancer subtyping. *BMC Bioinformatics*, 8:206, 2007.
16. A. I. Su, J. B. Welsh, L. M. Sapinoso, and *et al.* Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Research*, 61:7388–7393, 2001.
17. L. J. van 't Veer, H. Dai, M. J. van de Vijver, and *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415:530–536, 2002.
18. S. L. Pomeroy, P. Tamayo, M. Gaasenbeek, and *et al.* Prediction of central nervous system embryonal tumor outcome based on gene expression. *Nature*, 415:436–442, 2002.
19. D. Singh, P. G. Febbo, K. Ross, and *et al.* Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, 1:203–209, 2002.