International Journal of Pattern Recognition and Artificial Intelligence
Vol. 16, No. 7 (2002) 831–844

EXPLORING FEATURES AND CLASSIFIERS TO CLASSIFY GENE EXPRESSION PROFILES OF ACUTE LEUKEMIA*

SUNG-BAE CHO

Department of Computer Science, Yonsei University,
134 Shinchon-dong, Sudaemoon-ku, Seoul 120-749, Korea
sbcho@cs.yonsei.ac.kr

Bioinformatics has recently drawn a lot of attention to efficiently analyze biological genomic information with information technology, especially pattern recognition. In this paper, we attempt to explore extensive features and classifiers through a comparative study of the most promising feature selection methods and machine learning classifiers. The gene information from a patient's marrow expressed by DNA microarray, which is either the acute myeloid leukemia or acute lymphoblastic leukemia, is used to predict the cancer class. Pearson's and Spearman's correlation coefficients, Euclidean distance, cosine coefficient, information gain, mutual information and signal to noise ratio have been used for feature selection. Backpropagation neural network, self-organizing map, structure adaptive self-organizing map, support vector machine, inductive decision tree and k-nearest neighbor have been used for classification. Experimental results indicate that backpropagation neural network with Pearson's correlation coefficients produces the best result, 97.1% of recognition rate on the test data.

Keywords: Biological data mining; feature selection; classification; gene expression profile; acute leukemia; AML; ALL.

1. Introduction

There is no off-the-shelf way to cancer detection and class discovery right now, even though they have been seriously investigated over the past three decades. It is partially due to the fact that there can be so many pathways causing cancer, and there exist tremendous varieties. Recently, array technologies have made it easy to monitor the expression patterns of thousands of genes during cellular differentiation and response.[8,16] The gene expression data can be considered as just huge sequences of numbers, and the necessity of tools to analyze them to get useful information becomes radical.

Many researchers have been attempting to propose the optimal classification technique to work out this problem, especially dealing with predictive discrimination of Leukemia dataset as shown in Table 1.[3,5,7,11,14] Some produce better result

*This paper was supported by Brain Science and Engineering Research Program sponsored by the Korean Ministry of Science and Technology.

Table 1. Relevant works on Leukemia dataset.

Authors	Method		Accuracy [%]
	Feature	Classifier	
Furey *et al.*[3]	Signal to noise ratio	SVM	75.6
Li *et al.*[4]	Model selection with Akaike information criterion and Bayesian information criterion with logistic regression		94.1
Ben-Dor *et al.*[5]	All genes, TNoM score	Nearest neighbor	91.6
		SVM with quadratic kernel	94.4
		AdaBoost	95.8
Dudoit *et al.*[6]	The ratio of between-groups to within-groups sum of squares	Nearest neighbor	95.0 ∼
		Diagonal linear discriminant analysis	95.0 ∼
		BoostCART	95.0 ∼
Nguyen *et al.*[7]	Principal component analysis	Logistic discriminant	97.1
		Quadratic discriminant analysis	82.4
	Partial least square	Logistic discriminant	97.1
		Quadratic discriminant analysis	88.2

than others, but there have been still no comprehensive work to compare the possible methods with different feature selection and classification techniques. From the pattern recognition point of view, we need a thorough effort to give an evaluation of the possible methods to solve the problem of analyzing biological data.

In this paper, we attempt to explore the features and classifiers that efficiently detect the class of the cancer. To find out the genes that have cancer-related functions, we have adopted seven feature selection methods, which are commonly used in the field of data mining and pattern recognition: Pearson's and Spearman's correlations are based on a statistical approach, Euclidean distance and cosine coefficient based on similarity distance measure, and information gain and mutual information based on information theoretic approach. We have also utilized six classifiers: multilayer perceptron, self-organizing map (SOM), structure-adaptive SOM, support vector machine, decision tree and k-nearest neighbor.

2. DNA Microarray

DNA arrays consist of a large number of DNA molecules spotted in a systemic order on a solid substrate. Depending on the size of each DNA spot on the array, DNA arrays can be categorized as microarrays when the diameter of DNA spot is less than 250 microns, and macroarrays when the diameter is bigger than 300 microns. The arrays with the small solid substrate are also referred to as DNA chips. It is so powerful that we can investigate the gene information in short time, because at least hundreds of genes can be put on the DNA microarray to be analyzed.

DNA microarrays are composed of thousands of individual DNA sequences printed in a high density array on a glass microscope slide using a robotic arrayer as shown in Fig. 1. The relative abundance of these spotted DNA sequences in two DNA or RNA samples may be assessed by monitoring the differential hybridization

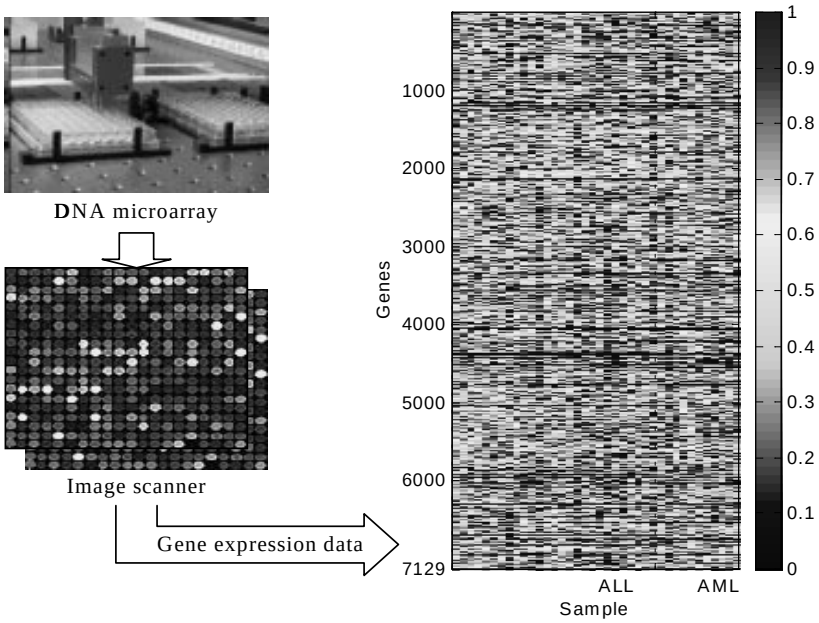

Fig. 1. General process of acquiring the gene expression data from DNA microarray.

of the two samples to the sequences on the array. For mRNA samples, the two samples are reverse-transcribed into cDNA, labeled using different fluorescent dyes mixed (red-fluorescent dye Cy5 and green-fluorescent dye Cy3). After the hybridization of these samples with the arrayed DNA probes, the slides are imaged using a scanner that makes fluorescence measurements for each dye. The log ratio between the two intensities of each dye is used as the gene expression data.[4,6,10]

$$\text{gene\_expression} = \log_2 \frac{\text{Int(Cy5)}}{\text{Int(Cy3)}} \tag{1}$$

where Int(Cy5) and Int(Cy3) are the intensities of red and green colors. Since at least hundreds of genes are put on the DNA microarray, it is so helpful that we can investigate the genome-wide information in short time.

3. Gene Expression Classification

We have used the data set of DNA microarray from the myeloid samples of the patients who have either the acute lymphoblastic leukemia (ALL) or the acute myeloid leukemia (AML). This data set is from Affymetrix arrays, not cDNA arrays that were explained in Sec. 2, but still it is one of the most well-known data sets. There are 7129 known genes in human genome, so that each sample has 7129 values of gene expressions. The system developed in this paper to predict the cancer class of patients is as shown in Fig. 2. After acquiring the gene expression data calculated

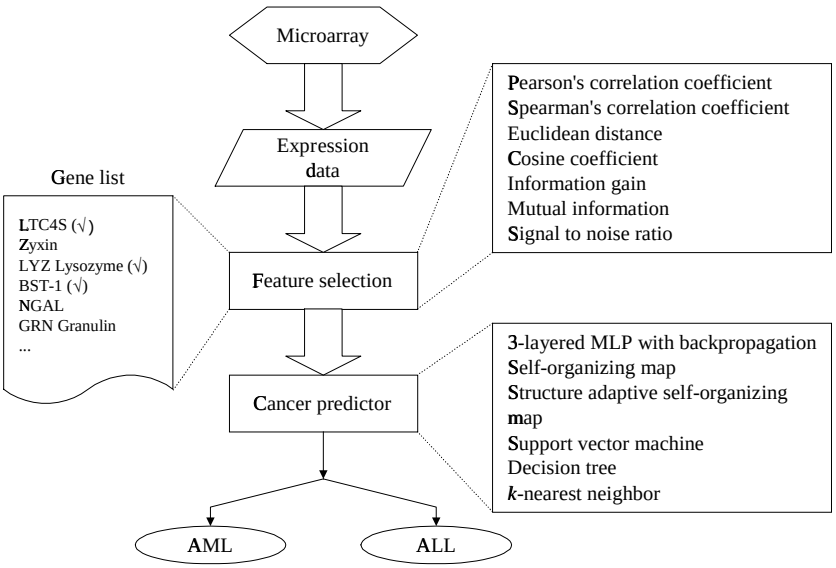

Fig. 2. Overview of cancer classification system.

from the DNA microarray, our prediction system goes through two stages: feature selection and pattern classification stages.

The feature selection can be thought of as the gene selection, which is to get the list of genes that might be informative for the prediction by statistical, information theoretical methods, etc. Since it is highly unlikely that all the 7129 genes have the information related to the cancer and using all the genes results in too big dimensionality, it is necessary to explore the efficient way to get the best feature. We have extracted 25 genes using seven methods described in Sec. 3.1, and the cancer predictor classifies the category only with these genes.

Given the gene list, a classifier makes the decision to which category the gene pattern belongs at prediction stage. We have adopted six most widely used classification methods as shown in Fig. 2.

3.1. *Features*

3.1.1. *Statistical approach*

Using the statistical correlation analysis, we can see the linear relationship and the direction of relation between two variables. Correlation coefficient r varies from -1 to $+1$, so that the data distributed near the line biased to $(+)$ direction will have positive coefficients, and the data near the line biased to $(-)$ direction will have negative coefficients.

Suppose that we have a gene expression pattern $\mathbf{g_i}(i = 1 \sim 7129)$. Each $\mathbf{g_i}$ is a vector of gene expression levels from 38 samples, $\mathbf{g_i} = (e_1, e_2, e_3, \ldots, e_{38})$. The first 27 elements $(e_1, e_2, \ldots, e_{27})$ are examples of ALL, and the other 11 $(e_{28}, e_{29}, \ldots, e_{38})$

are those from AML. An ideal gene pattern that belongs to ALL class is defined by $g_{\text{ideal\_ALL}} = (1, 1, \ldots, 1, 0, \ldots, 0)$, so that all the elements from ALL samples are 1 and the others are 0. In this paper, we have calculated the correlation coefficient between this g_{ideal} and the expression pattern of each gene. When we have two vectors $\mathbf{X}$ and $\mathbf{Y}$ that contain N elements, r_{Pearson} and r_{Spearman} are calculated as follows:

$$r_{\text{Pearson}} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}} \tag{2}$$

$$r_{\text{Spearman}} = 1 - \frac{6\sum (D_x - D_y)^2}{N(N^2 - 1)} \tag{3}$$

where D_x and D_y are the rank matrices of $\mathbf{X}$ and $\mathbf{Y}$, respectively.

3.1.2. *Distance measure approach*

The similarity between two input vectors $\mathbf{X}$ and $\mathbf{Y}$ can be thought of as distance. Distance is a measure on how far the two vectors are located, and the distance between $g_{\text{ideal\_ALL}}$ and g_i tells us how much the g_i is likely to be in the ALL class. Calculating the distance between them, if it is bigger than a certain threshold, the gene g_i would belong to ALL, otherwise g_i belongs to AML. In this paper, we have adopted Euclidean distance ($r_{\text{Euclidean}}$) and cosine coefficient (r_{Cosine}) represented by the following equations:

$$r_{\text{Euclidean}} = \sqrt{\sum (X - Y)^2} \tag{4}$$

$$r_{\text{Cosine}} = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}. \tag{5}$$

3.1.3. *Information-theoretic approach*

We have utilized the information gain and mutual information that are widely used in many fields such as text categorization and data mining. If we count the number of genes excited ($P(g_i)$) or not excited ($P(\bar{g}_i)$) in category $c_j(P(c_j))$, the coefficients of the information gain and mutual information become as follows:

$$IG(g_i, c_j) = P(g_i|c_j) \log \frac{P(g_i|c_j)}{P(c_j) \cdot P(g_i)} + P(\bar{g}_i|c_i) \log \frac{P(\bar{g}_i|c_j)}{P(c_j) \cdot P(\bar{g}_i)} \tag{6}$$

$$MI(g_i, c_j) = \log \frac{P(g_i, c_j)}{P(c_j) \cdot P(g_i)}. \tag{7}$$

Mutual information tells us the dependency relationship between two probabilistic variables of events. If two events are completely independent, the mutual information is 0. The more they are related, the higher the mutual information. Information

gain is used when the features of samples are extracted by inducing the relationship between gene and class by the presence frequency of the gene in the sample. Information gain measures the goodness of gene using the presence and absence within the corresponding class.

For each gene g_i, some are from ALL, and some are from AML samples. If we calculate the mean μ and standard deviation σ from the distribution of gene expressions within their classes, the signal to noise ratio of gene g_i, $SN(g_i)$, is defined by:

$$SN(g_i) = \frac{\mu_{\mathrm{ALL}}(g_i) - \mu_{\mathrm{AML}}(g_i)}{\sigma_{\mathrm{ALL}}(g_i) - \sigma_{\mathrm{AML}}(g_i)}\,. \tag{8}$$

3.2. *Classifiers*

3.2.1. *Multilayer perceptron*

Error backpropagation neural network is a feed-forward multilayer perceptron (MLP) that is applied in many fields due to its powerful and stable learning algorithm.[12] The neural network learns the training examples by adjusting the synaptic weight of neurons according to the error occurring on the output layer. The power of the backpropagation algorithm lies within two main aspects: local for updating the synaptic weights and biases, and efficient for computing all the partial derivatives of the cost function with respect to these free parameters.[2]

3.2.2. *Self-organizing map*

Self-organizing map (SOM) defines a mapping from the input space onto an output layer by unsupervised learning algorithm. SOM has an output layer consisting of N nodes, each of which represents a vector that has the same dimension as the input pattern. For a given input vector X, the winner node m_c is chosen using Euclidean distance between $\boldsymbol{x}$ and its neighbors, $\boldsymbol{m_i}$.

$$\|x - m_c\| = \min_i \|x - m_i\| \tag{9}$$

$$m_i(t+1) = m_i(t) + \alpha(t) \times n_{ci}(t) \times \{x(t) - m_i(t)\}\,. \tag{10}$$

Here, $\alpha(t)$ is the learning rate and $n_{ci}(t)$ is a neighborhood function.

3.2.3. *Structure-adaptive SOM*

Even though SOM is well known for its good performance of topology preserving, it is difficult to apply it to practical classification since the topology should be fixed before training. A structure adaptive self-organizing map (SASOM) is proposed to overcome this shortcoming.[9] SASOM starts with 4×4 map, and dynamically splits the output nodes of the map, where the data from different classes are mixed, trained with the LVQ learning algorithm. Figure 3 illustrates the algorithm of SASOM.

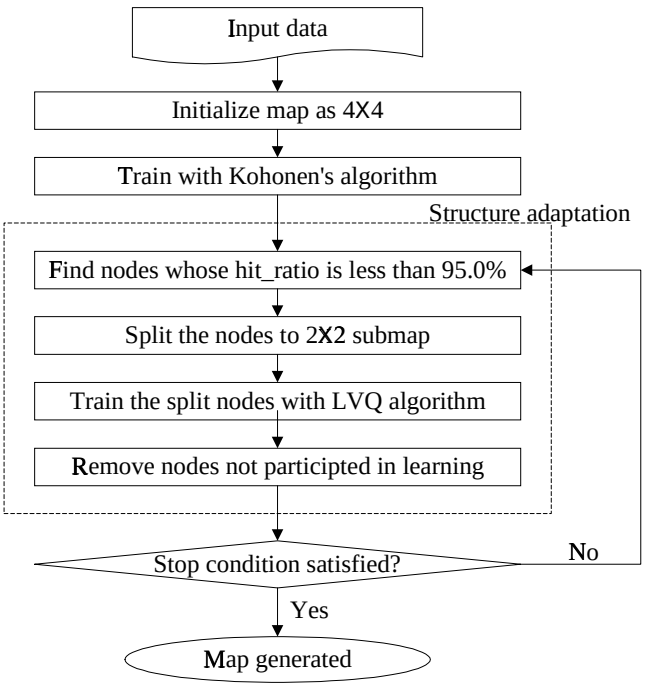

Fig. 3. Flow chart of structure-adaptive self-organizing map.

3.2.4. *Support vector machine*

Support vector machine (SVM) estimates the function classifying the data into two classes.[13,17] SVM builds up a hyperplane as the decision surface in such a way to maximize the margin of separation between positive and negative examples. SVM achieves this by the structural risk minimization principle that the error rate of a learning machine on the test data is bounded by the sum of the training-error rate and a term that depends on the Vapnik–Chervonenkis (VC) dimension. Given a labeled set of M training samples $(\boldsymbol{X_i}, Y_i)$, where $\boldsymbol{X_i} \in R^N$ and Y_i is the associated label, $Y_i \in \{-1, 1\}$, the discriminant hyperplane is defined by:

$$f(x) = \sum_{i=1}^{M} Y_i \alpha_i k(X, X_i) + b \tag{11}$$

where $k(.)$ is a kernel function and the sign of $f(X)$ determines the membership of X. Constructing an optimal hyperplane is equivalent to finding all the nonzero α_i (support vectors) and a bias b. We have used SVM^{light} module in this paper.

3.2.5. *Decision tree*

The concept-learning induction method such as decision tree (DT) aims to construct rules for the classification from the set of objects of which class labels are

known. Quinlan's C4.5 uses an information-theoretical approach based on the energy entropy. C4.5 builds the decision tree as follows: select an attribute, divide the training set into subsets characterized by the possible values of the attribute, and follow the same partitioning procedure recursively with each subset until no subset contains objects from more than one class. The single class subsets correspond to the leaves. The entropy-based criterion that has been used for the selection of the attribute is called the gain ratio criterion.[15]

Let $\mathbf{X}$ be a possible test (attribute selection) that partitions the training set $\mathbf{T}$ into n subsets $(T_1, T_2, \ldots, T_n)$, split_info($\mathbf{X}$), as the entropy of a message where information is given in terms of outcomes, and gain_ratio($\mathbf{X}$) can be defined as follows:

$$\text{split\_info}(X, T) = -\sum \left(\frac{|T_i|}{|T|}\right) \log_2 \left(\frac{|T_i|}{|T|}\right) \tag{12}$$

$$\text{gain\_ratio}(X) = \frac{\text{gain}(X)}{\text{split\_info}(X)}\,. \tag{13}$$

The gain ratio criterion selects the test $\mathbf{X}$ so that the gain_ratio($\mathbf{X}$) is maximized.

3.2.6. *k-nearest neighbor*

k-nearest neighbor (KNN) is one of the most common methods among memory based induction. Given an input vector, KNN extracts k closest vectors in the reference set based on similarity measures, and takes the decision for the label of input vector using the labels of the k nearest neighbors.

Pearson's correlation has been used as the similarity measure. When we have an input X and a reference set $D = d_1, d_2, \ldots, d_N$, the probability that X may belong to class c_j, $P(X, c_j)$ is defined as follows:

$$P(X, c_j) = \sum_{d_i \in kNN} \text{Sim}(X, d_i) P(d_i, c_j) - b_j \tag{14}$$

where $\text{Sim}(X, d_i)$ is the similarity between X and d_i, and b_j is a bias term.

4. Experimental Results

4.1. *Environments*

We have used a collection of expression measurements provided by Golub *et al.*[8] Among 72 bone marrow samples obtained from acute leukemia patients, 38 samples were used for training and the other 34 were for test. The training data consists of 27 ALL and 11 AML samples, whereas the test data has 20 ALL and 14 AML samples. ALL class is encodes as 1, and AML class as 0. Each sample consists of 7129 gene expression profiles.

For feature selection, each gene is scored based on the feature selection methods described in Sec. 3.1, and the 25 top-ranked genes are chosen as the feature of

```
   ID  Name
   ====+=========================================================================
   3320  Leukotriene C4 synthase (LTC4S) gene
   2020  FAH Fumarylacetoacetate
   1745  LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog
   5039  LEPR Leptin receptor
   4196  "PRG1 Proteoglycan 1, secretory granule"
   2288  DF D component of complement (adipsin)
   6201  INTERLEUKIN-8 PRECURSOR
   1882  CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)
   2121  CTSD Cathepsin D (lysosomal aspartyl protease)
   6200  Interleukin 8 (IL8) gene
   2043  "LGALS3 Lectin, galactoside-binding, soluble, 3 (galectin 3) (NOTE: redefinition of
symbol)"
```

Fig. 4. Genes selected by Pearson's correlation.

the input pattern. For classification, we have used three-layered MLP with $5 \sim 15$ hidden nodes, two output nodes, $0.03 \sim 0.5$ of learning rate and $0.1 \sim 0.9$ of momentum. SOM is used by $2 \times 2 \sim 5 \times 5$ map with rectangular topology and 0.05 of learning rate. KNN is used with $k = 1 \sim 38$. The best parameters have been chosen after a thorough empirical study. The final results are averaged by 10 runs of experiments.

4.2. *Analysis of results*

Figure 4 shows the Ids and names of 11 out of 25 genes chosen by Pearson's method, and also selected by the other three feature selection methods. Eight out of them appear in the result of cosine coefficient method, four in Spearman's correlation and three in information gain. g_{2288} has appeared four times, and g_{6200} has been top-ranked both in information gain and Spearman's correlation, which imply they are very informative. There are no genes that appear in every method simultaneously.

The results of the recognition rate on the test data are as shown in Table 2. The MLP seems to produce the best recognition rate among the classifiers on the average. SASOM performs better than SOM, and DT has good results in some cases, but it seems to be very dependent on the features used. KNN does not seem to perform the classification at all. In the meanwhile, Pearson's and cosine coefficients have better results than the other feature selection methods, obtaining 87% on

Table 2. Recognition rate by features and classifiers (%).

Feature/Classifier	MLP	SOM	SASOM	SVM	DT	KNN ($k = 10$)
Pearson Coefficient	97.1	73.5	88.2	79.4	97.1	29.4
Spearman Coefficient	70.6	73.5	82.4	88.2	82.4	32.4
Euclidean Distance	97.1	70.6	70.6	58.5	73.5	32.4
Cosine Coefficient	79.4	97.1	74.1	94.1	94.1	23.5
Information Gain	91.2	73.5	82.4	88.2	82.4	58.8
Mutual Information	67.6	67.6	64.7	58.5	47.1	58.8
S/N Ratio	94.1	52.9	64.7	58.5	55.9	8.8

the average (except KNN results). Information gain has 83.5% and Spearman's correlation has 79.4% of recognition rate. This fact indicates that several genes are chosen in common from those four feature selection methods, which means they may be correlated somehow.

Figures 5 and 6 show the comparison of the average performance with respect to the features and the classifiers, respectively. Information gain is the best, Pearson's correlation is the second, and signal to noise ratio is the worst. On the other hand, MLP is the best, whereas k-nearest neighbor is the worst. From this analysis, we can expect the best performance with the MLP with the feature from information gain or Pearson's correlation.

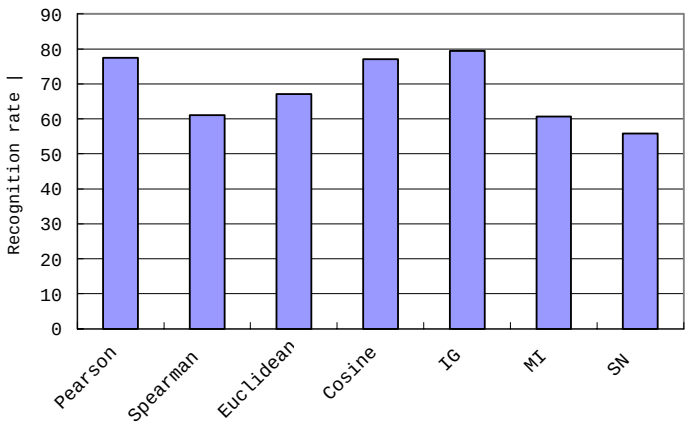

Fig. 5. Comparison of average performance of feature selection methods.

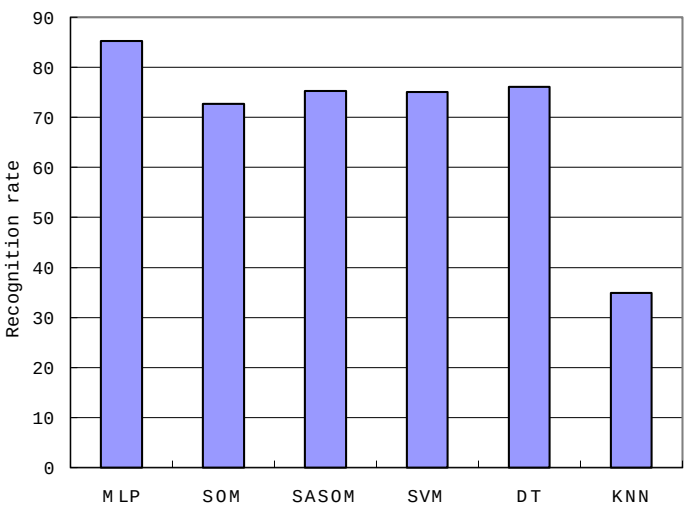

Fig. 6. Comparison of average performance of classification methods.

Feature	Multilayer perceptron	Structure-adaptive SOM	Decision tree
Pearson	**28**	**21 23 27 28**	**11**
Spearman	**15 16 20 22 23 24 25 27 28 29**	**14 17 21 24 25 28**	**5 6 9 17 21 28**
Euclidean	**22**	**11 14 16 19 22 23 24 25 26 28**	**16 19 20 22 23 24 25 26 28**
Cosine	**11 16 22 24 25 26 28**	**23 29**	**23 28**
IG	**16 22 26**	**5 6 9 17 21 28**	**14 17 21 24 25 28**
MI	**4 9 12 16 18 19 22 24 25 26 31**	**2 13 16 19 20 22 23 24 26 28 29 33**	**1 2 3 5 6 9 13 14 16 17 18 19 21 22 23 24 26 27**
SN	**28 29**	**9 11 12 14 16 21 22 23 24 27 28 32**	**1 4 5 6 9 10 12 17 18 19 21 28 29 30 33**

Fig. 7. Misclassified samples.

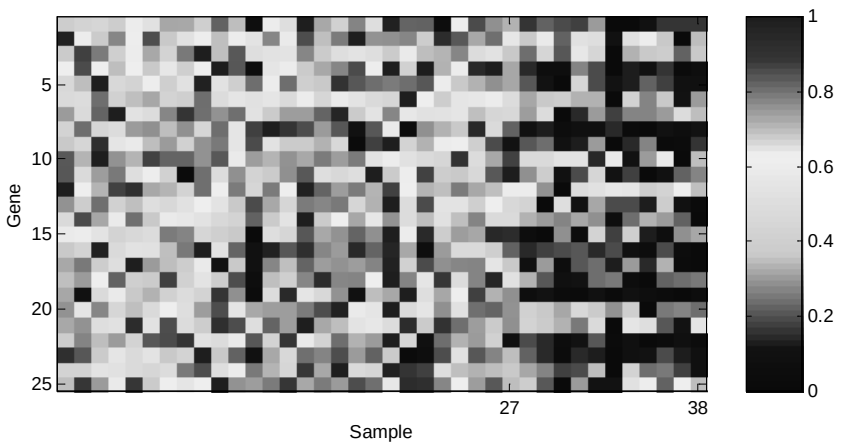

Fig. 8. Expression level of genes chosen by r_{Pearson}.

Figure 7 shows the examples of misclassification made by MLP, SASOM and DT. Sample 28 is the only one misclassified by MLP with the features selected by Pearson's coefficients. Many other classifiers also fail to classify this sample correctly, but MLP with Euclidean distance, information gain and mutual information, SASOM with cosine, and decision tree with Pearson's correlation and mutual information features have qualified as the correct classification. On the other hand, samples 11 and 29, which are misclassified by MLP with signal-to-noise ratio and decision tree with Pearson's correlation respectively, have been correctly classified by most of other classifiers. Figure 8 shows the expression level of genes chosen by Pearson's correlation method.

5. Concluding Remarks

It is well known that distinguishing acute leukemias is one of the challenging problems in bioinformatics because the appearance is highly similar, and there have been many attempts to work out this problem with different features and classification methods as shown in Table 1. We have conducted a thorough quantitative

comparison among the 42 combinations of features and classifiers. Pearson's correlation, Spearman's correlation, cosine coefficient and information gain are the top four feature selection methods, and MLP, SASOM and decision tree are the best classifiers. The experimental results also imply some correlations between features and classifiers, which might guide the researchers to choose or devise the best classification method for their problems in bioinformatics. Based on the results, we are undergoing to develop the optimal feature/classifier combination to produce the best performance on the classification of gene expression profiles for acute leukemia.

Moreover, the data set used is too small, and there are many benchmarks in this field. Some of them are lymphoma data (http://llmpp.nih.gov/lymphoma/) and colon data. Lymphoma data contain 4026 genes across 47 samples, of which 24 are referred to as germinal center B-like DLBCL and 23 as activated B-like DLBCL. The original colon data contain the expression levels of 2000 genes across 62 samples, of which 40 are tumor tissue and 22 are normal tissue.[1] We will have to conduct the same experiments with those data sets to strengthen the results obtained.

References

1. U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack and A. J. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proc. Natl. Acad. Sci. USA* **96** (1999) 6745–6750.
2. H. D. Beale, *Neural Network Design*, PWS Publish Company, Chap. 11, pp. 1–47, 1996.
3. A. Ben-Dor, L. Bruhn, N. Friedman, I. Nachman, M. Schummer and N. Yakhini, "Tissue classification with gene expression profiles," *J. Comput. Biol.* **7** (2000) 559–584.
4. J. Derisi, V. Iyer and P. Brosn, "Exploring the metabolic and genetic control of gene expression on a genomic scale," *Science* **278** (1997) 680–686.
5. S. Dudoit, J. Fridlyand and T. P. Speed, "Comparison of discrimination methods for the classification of tumors using gene expression data," Technical Report 576, Department of Statistics, University of California, Berkeley, June 2000.
6. M. Eisen, P. Spellman, P. Brown and D. Bostein, "Cluster analysis and display of genome-wide expression patterns," *Proc. Natl. Acad. Sci. USA* **95** (1998) 14863–14868.
7. T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics* **16**, 10 (2000) 906–914.
8. 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 and E. S. Lander, "Molecular classification of cancer: class discovery and class prediction by gene-expression monitoring," *Science* **286** (1999) 531–537.
9. H. D. Kim and S.-B. Cho, "Genetic optimization of structure-adaptive self-organizing map for efficient classification," *Proc. Int. Conf. Soft Computing*, World-Scientific Publishing, October 2000, pp. 34–39.

10. D. Lashkari, J. Derisi, J. McCusker, A. Namath, C. Gentile, S. Hwang, P. Brown and R. Davis, "Yeast microarrays for genome wide parallel genetic and gene expression analysis," *Proc. Natl. Acad. Sci. USA* **94** (1997) 13057–13062.
11. W. Li and Y. Yang, "How many genes are needed for a discriminant microarray data analysis," *Critical Assessment of Techniques for Microarray Data Mining Workshop*, December 2000.
12. R. P. Lippman, "An introduction to computing with neural nets," *IEEE ASSP Mag.* April (1987) 4–22.
13. B. Moghaddam and M.-H. Yang, "Gender classification with support vector machines," *Proc. 4th IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2000, pp. 306–311.
14. D. V. Nguyen and D. M Rocke, "Tumor classification by partial least squares using microarray gene expression data," *Bioinformatics* **18**, 1 (2002) 39–50.
15. J. R. Quinlan, "The effect of noise on concept learning," *Machine Learning*: *An Artificial Intelligence Approach*, eds. R. S. Michalski, J. G. Carbonell and T. M. Mitchell, Morgan Kauffmann, San Mateo, CA, Vol. 2, 1986, pp. 149–166.
16. P. Tamayo, "Interpreting patterns of gene expression with self-organizing map: Methods and application to hematopoietic differentiation," *Proc. Natl. Acad. Sci.* **96** (1999) 2907–2912.
17. V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, NY, 1995.

Sung-Bae Cho received the B.S. degree in computer science from Yonsei University, Seoul, Korea, in 1988 and the M.S. and Ph.D. degrees also in computer science from KAIST (Korea Advanced Institute of Science and Technology), Taejeon, Korea, in 1990 and 1993, respectively.

He worked as a Member of the Research Staff 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.

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 is 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.

His research interests include neural networks, pattern recognition, intelligent man-machine interfaces, evolutionary computation and artificial life.