

CANCER CLASSIFICATION BASED ON MICROARRAY DATA USING BINARY-TUBE TREE

Suebkul Kanchanasuk and Krung Sinapiromsaran*

Received: September 09, 2018; Revised: June 28, 2019; Accepted: June 28, 2019

Abstract

Analysis of large gene expression data via machine learning has been widely acknowledged to accurately predict cancer classes from the microarray data which are cast as the numerical matrix from a collection of complex RNAs. This research proposes a novel continuous value binary-tube tree constructed from the core vector of multi-features and the tolerance radius of the specific cancer types. To validate this classifier, the accuracy measure and Matthews correlation coefficient measure are compared with the support vector machine and the decision tree classifier using k-folds cross-validation. The result shows that the binary-tube tree outperforms the support vector machine and the decision tree classifier for leukemia, colon cancer, lymphoma, and breast cancer; only the lung cancer data set shows no significant improvement.

Keywords: Classification, microarray data, binary-tube tree, decision tree

Introduction

In the past decade, microarray gene expression data has been an interesting issue in microarray is one of the technologies which facilitates many researchers to analyze and uncover issues which have never been revealed before. This technique can analyze the expression of several thousand genes in a single reaction simultaneously. Due to a high-throughput molecular method, mining tools to extract the knowledge and also analyze the data are

required. In a clinical trial, the DNA microarray can be expressed to (1) confirm the diagnostics results, (2) classify the disease, (3) monitor the response to therapy, and (4) identify the evolutionary mechanisms of the disease (Tarca *et al.*, 2006). Clinical research tools using machine learning approaches in DNA microarray data are considered as significant advances in the field. Zhao *et al.* (2007) investigated the co-regulated gene in

Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. E-mail: krung.s@chula.ac.th; suebkul.k@gmail.com

** Corresponding author*

microarray chips using the G-cluster model to reveal the positive and negative co-regulated genes. One of the most cited literature works is Golub (1999) who used the feature selection and nearest neighbor algorithm to discover a class of gene expression from acute leukemia patients. Moreover, the hierarchical clustering algorithm was adopted to classify the tumor types of DNA microarray data sets, such as lymphomas by and breast cancer by van't Veer *et al.* (2002). In addition, Nguyen and Rocke (2002) designed a classification algorithm using the partial least square method to reduce the number of genes (features) and used the linear discriminant and quadratic discriminant analysis to classify 5 tumor data sets: ovarian, leukemia, lymphoma, colon, and lung cancer. Rafii *et al.* (2016) used the ART2 network to diagnose the DNA microarray data set. One of the favorite classification algorithms for DNA microarray data sets is the support vector machine (SVM) (Furey *et al.*, 2000; Cho and Won, 2003; Mukherjee, 2003; Mohamad and Deris, 2005; Lewis *et al.*, 2006; Bennet *et al.*, 2015). The SVM algorithm is a supervised

machine learning algorithm which determines the optimal hyperplane that distinguishes between 2 target classes. The application of the SVM to DNA microarray data sets demonstrates its effectiveness from informative features or attributes such as significantly correlated genes. The SVM requires an appropriate setting of parameters (cost, type of the kernel function, etc.) to achieve the intended performance. The favorite machine learning algorithm, because it is reliable and has the simplest interpretation, is the decision tree classifier. This classifier uses measurement to evaluate the rules or the patterns from the data. The criteria measurements for the decision tree are selected from various procedures.

Figure 1 shows the rules of the specific algorithms to classify the data. The SVM specifies that the data uses the hyperplane as shown in Figure 1(a) and the decision tree in Figure 1(b) is specified by the splitting criteria which are computed by the information gain. In general, a DNA microarray data set consists of a large number of genes (features), whereas many machine learning algorithms are not originally designed with linear discriminant and quadratic discriminant analysis to classify 5 tumor data sets to cope with such numerous features. Hence, much of the research has incorporated only the correlated genes to extract the suitable features for working with the well-known machine learning algorithm.

In this research, the classification algorithm, called the binary-tube tree (BITT) algorithm, has been proposed to classify specific types of cancers from the DNA microarray data set. The BITT algorithm, which is modified from the recursive binary-tube partitioning (RBTP) algorithm (Kanchanasuk and Sinapiromsaran, 2016), is a binary classifier based on the decision tree concept to find the combination of the expression levels of genes which is relevant to a specific tumor type or cell type in terms of a “tube”, as shown in Figure 3.

A tube is designed to capture the characteristics of the gene expression of the target class by identifying the several tube components. A tube consists of 3 components, a core vector, a tube length, and a tube radius where the core vector is denoted as the

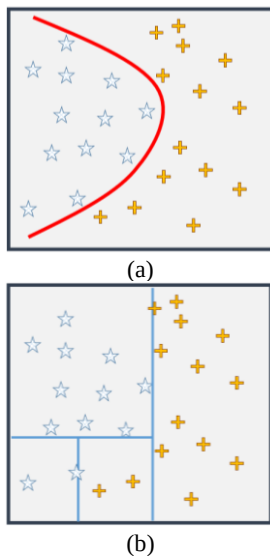


Figure 1. Machine learning algorithm paradigm: (a) support vector machine using radial basis kernel function and (b) a decision tree classifier using information gain

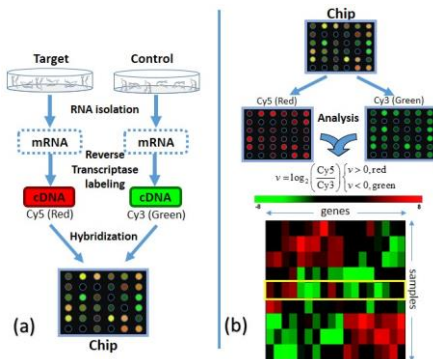


Figure 2. (a) A typical DNA microarray hybridization experiment and (b) quantitative analysis of 2 intensity levels from transformed log₂ ratio samples and rearrangement of the samples to the row of the matrix and of the genes to the column

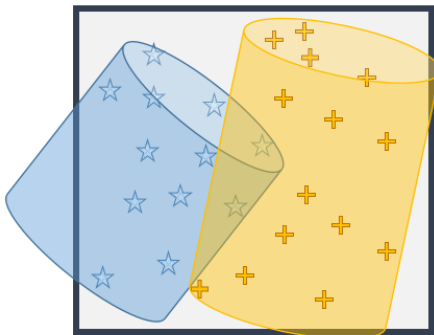


Figure 3. Visualization of the BITT algorithm

directional stretch of gene expression of the samples, a tube length is defined as the length of the stretch, and the tube radius is the orthogonal stretch from the core vector.

The concept of the BITT algorithm remedies the weakness of a decision tree by using a tube instead of the best selected attribute. Normally, samples in 1 cancer type tend to form a cluster so that the centroid is a good representation of this multivariate data. The sample that is situated furthest from this centroid also demonstrates the longest tail of the cluster which can be captured by the core vector and tube length.

Support Vector Machine

The SVM was first introduced as a supervised learning model by Cortes and Vapnik (1995) for binary classification. The SVM was successfully used in several bioinformatics applications, such as oncology with magnetic resonance imaging for electroencephalography (Lotte *et al.*, 2018), and molecular classification in a DNA microarray. The SVM is a learning kernel-based system that splits a high dimensional feature space using a hyperplane as the linear separator. Unlike an artificial neural network which tries to define a complex function in the input feature space, the SVM uses the kernel method to capture a nonlinear mapping of the complex data in high dimensional feature spaces and then uses a simple linear function to create a linear decision boundary. Thus, the problem of gene expression profiling can be cast as the problem of choosing a suitable kernel for the data projection with a hyperplane separation. The parameters of an SVM model are determined based on risk minimization to search for the target partition known as the optimal hyperplane. Although the optimal hyperplane, which is applied as the kernel function, is used as the criterion for classifying the samples, conflict regions occur from misclassification of the class of the samples during the training phase.

Decision Tree

A decision tree consists of a set of nodes and a set of edges. A node is categorized as an internal node or a leaf node where the root node is just the first internal node. All internal nodes represent the split criteria to partition a data set into various subsets based on a single characteristic. An edge represents 1 of the possible outcomes of the test condition from the internal node. Finally, a leaf node is labeled by the class. In the past decade, 1 of the top 10 algorithms in data mining is C4.5 (Wu and Kumar, 2009) which is commonly used for building a decision tree. C4.5, introduced by Quinlan (1993), uses the gain ratio to measure

the branch on the feature which returns the maximum information gain ratio. ID3 and C4.5 use the entropy measurement to find the best split from the feature having the highest information gain, while C4.5 also uses Shannon entropy as an impurity measure.

Let $Entropy(D)$ be an overall impurity of the data set D computed by:

$$Entropy(D) = - \sum_{i=1}^{|C|} p_i \log p_i \quad (1)$$

where p_i is the approximated probability of the number of instances in class i and the set of classes is denoted by C . Moreover, let the additional measurement:

$$Entropy_v(D) = \sum_{v \in V} \frac{|D_v|}{|D|} Entropy(D_v) \quad (2)$$

be the entropy of the subsets of the data set D with the value v belonging to the feature V . The information gain is:

$$InfoGain(V) = Entropy(D) - Entropy_v(D). \quad (3)$$

where V is the selected feature. The split information measure, which is computed by:

$$SplitInfo_v(D) = - \sum_{v \in V} p_v \log p_v, \quad (4)$$

shows the splitting value v , where p_v is the probability of an instance with the attribute V having the value v . Furthermore, C4.5 deals with this problem using a gain ratio calculated by:

$$GainRatio = InfoGain(V) / SplitInfo_v(D). \quad (5)$$

The limitation of a decision tree is due to its simplicity of using 1 attribute selection in each step, which may not help to split clustering instances along the non-orthogonal dimensions.

Materials and Methods

Binary-Tube Tree (BITT)

The RBTP algorithm (Kanchanasuk and Sinapiromsaran, 2016) is modified for use as the BITT algorithm to classify the DNA microarray data. For a binary class problem, the RBTP algorithm generates a classification-decision tree for the given training data set by recursively partitioning through the use of a tube. Each internal node (including the root node) has 2 tubes as a decision rule and the leaf node is a single class labeled node. A tube has 3 components, a core vector, a tube length, and a tube radius which is constructed by the same class samples.

The BITT algorithm is the binary classification algorithm for the gene expression profiling which modifies the core vector of the RBTP algorithm using the centroidal extreme instead of 2 extreme poles. The centroidal extreme is defined as the farthest sample from the centroid of all the samples.

The implementation of the BITT algorithm has been described as follows. The first step performs the data partition using the target classes which will be used to generate a tube of each class. A single tube consists of 3 components: a core vector ($\overline{pc}$) which is a vector with the direction starting from the centroidal extreme (p) to the centroid (c), the magnitude of the core vector which is the tube length (H), and the maximum extent along the core vector which is the tube radius (R).

The details of building a tube are as follows. Given a set of gene expression samples (X) with s samples, N is the number of features in the experiment, and the cancer types in the experiment (Y) which can be labeled as $Y = \{type_1, type_2\}$. Let $x^i = (x^i_1, x^i_2, x^i_3, \dots, x^i_N)$ be a tuple of the expression levels of sample i . In this paper, the Euclidean distance is defined as a distance measure, so

$d(x^i, x^j) = \sqrt{\sum_{m=1}^N (x^i_m - x^j_m)^2}$ the distance between two samples x^i and x^j . Moreover, the magnitude of a vector is defined by $\|\overline{ab}\| = d(a, b)$. The

centroid can be computed by $c = \frac{1}{S} \sum_{i=1}^S x^i$. The

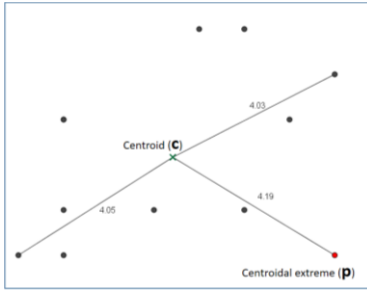


Figure 4. The centroid and the centroidal extreme

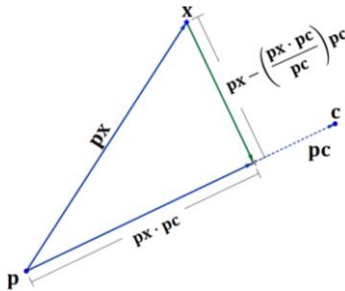


Figure 5. Example of a tube in 2D view

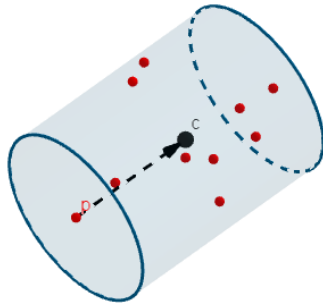


Figure 6. Visualization of a tube in 3D view

centroidal extreme (p) is the sample in X which is the furthest distance to the centroid c , so $p = \arg \max_{x \in X} \{d(x, c)\}$. The core vector of the BITT algorithm is a unit vector from the centroidal extreme to the centroid of the data set, defined by $\overline{pc} = \frac{pc}{\|pc\|}$.

The tube length is the maximum magnitude of the projection vector of $\overline{px}$ onto

the core vector, computed by $H = \arg \max \{\overline{px} \cdot \overline{pc}\}$.

The tube radius is the maximum of the shortest distance of X to $\overline{pc}$, calculated by $R = \arg \max_{x \in X} \left\{ \left\| \overline{px} - (\overline{px} \cdot \overline{pc}) \overline{pc} \right\| \right\}$. The illustrations of a tube in 2D and 3D are shown in Figures 5 and 6, respectively.

When the process to generate the binary tube from the samples of each class in the first step is applied, the samples are covered by the tube which is specified by the individual class. Samples that appear in the unique class tube will be classified as that class. So the recursive partitioning step will be applied to samples appearing in more than 1 tube. This will stop when all samples can be classified as 1 class or the number of samples is too small to generate a tube.

Performance Measures and Models Evaluation

Performance Measures

In this research, the performance of the classifiers is evaluated by 2 performance measures, accuracy and the Matthews correlation coefficient (MCC) using the confusion matrix of a binary class. The confusion matrix of a binary class is shown in Table 1.

The accuracy measure is the proportion of the true classified results (TP and TN) among the total number of samples (TP, TN, FP, and FN), calculated by:

$$ACCURACY = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

The MCC (Matthews, 1975) is the evaluating algorithm measure which is suitable for the unbalanced class of the data set. The formula of the MCC is calculated by values in the 2×2 confusion matrix, computed by:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (7)$$

Table 1. Confusion matrix for binary classification

		Predict Condition	
		Positive	Negative
Actual Condition	Positive	True Positive (TP)	False Negative (FN)
		False Positive (FP)	True Negative (TN)
	Negative	False Positive (FP)	True Negative (TN)
		True Positive (TP)	False Negative (FN)

because the MCC value has an equal value if any binary-class is defined as a positive class. This is an advantageous concept for the data set and is required to indicate the positive class of the data set. The MCC value is between -1 to 1. A coefficient of +1 indicates a perfect classification and -1 indicates a total disagreement of the classification.

Moreover, the 2 performance measures were tested by the non-parametric statistical test using the Wilcoxon signed-rank test to confirm the significance of the performance for the right-sided hypothesis testing with $\alpha = 0.05$.

Models Evaluation

The Wilcoxon signed-rank test, which was proposed by Wilcoxon (1945), is a non-parametric statistical test to compare 2 experimental results in the same sample group. This test is used in this research to rank the differences of the performance measure between 2 classifiers. The concept of this test is to assess the paired score differences and rank them in an absolute value from the win and loss classification. The better classifier shows that the minimum sum of the winning rank or losing rank is close to 0. The method ranks the differences in performance between 2 classifiers. This test is recommended when the data set is not a normal distribution from the least effect on the test.

Data Sets

There are 5 data sets from different gene expression databases which are described in Table 2 and are used in this paper. The first column is the name of the data set, the second column shows the class labels, the third

column shows the number of samples, while the last column contains the number of genes in the experiment. The microarray databases in this experimentation are cast as the binary classification problem. The details of the DNA microarray data sets which are shown in Table 2 were selected from the following sources.

The *leukemia data set* was published by Golub (1999). The data set consists of the gene expression of 7,130 genes and 72 samples. The class of samples is the collection of acute leukemia cells, acute myeloid leukemia (AML), and acute lymphocytic leukemia (ALL).

The *colon cancer data set* was collected by Alon *et al.* (1999) containing the gene expression of 40 tumors and 22 normal colon tissue samples. It was analyzed with an Affymetrix oligonucleotide array complementary that generates more than 6,500 human genes. The preprocessing step was applied to reduce the number of informative genes to 2,001 genes and label the type of the samples as a class with the normal tissue samples as “positive” and the tumor colon tissue samples as “negative”.

The *lymphoma data set* (Shipp *et al.*, 2002) consists of gene expression levels measured from Affymetrix chips. Each oligonucleotide array contained 7,129 genes sets from 77 samples with 58 samples of diffuse large B-cell lymphoma (DLBCL) and 19 samples of follicular lymphoma (FL).

The *lung cancer data set* which was collected by Gordon *et al.* (2002), contains 181 tissue samples from 12,533 probes with 150 samples of adenocarcinoma (ADCA), and 31 samples of malignant pleural mesothelioma (MPM).

The *breast cancer data set* was modified by Hoshida *et al.* (2007) from van't Veer *et al.* (2002). This data set consists of 3 classes (type 1: T1, type 2: T2, and type 3: T3) of 98 samples having 1,213 genes to be analyzed.

In this research, the experimental framework was designed based on Hong and Cho (2006) and Lewis *et al.* (2006) to test the performance of the classifiers using 5-fold cross-validation. The performance measures in this research consist of the accuracy measure and the MCC measure. Moreover, this

Table 2. Summary of the DNA microarray data set used in the experiments

Data set	Class			Samples	Genes
Leukemia	ALL (47)	AML (25)		72	7,130
Colon	Negative (40)	Positive (22)		62	2,001
Lymphoma	DLBCL (58)	FL (19)		77	7,129
Lung	ADCA (150)	MPM (31)		181	12,533
Breast cancer	Type1 (11)	Type2 (51)	Type3 (36)	98	1,213

Table 3. Predictive accuracy of C4.5, SVM, and BITT algorithm

Data set	Accuracy		
	C4.5	SVM	BITT
Breast (T1 vs. T2)	*0.862±0.03	*0.836±0.01	0.946±0.01
Breast (T1 vs. T3)	*0.874±0.03	*0.903±0.01	0.966±0.01
Breast (T2 vs. T3)	*0.936±0.03	*0.961±0.01	0.973±0.01
Colon	*0.793±0.03	*0.778±0.02	0.851±0.01
Leukemia	0.928±0.03	0.934±0.01	0.958±0.01
Lung	0.933±0.01	0.971±0.00	0.944±0.01
Lymphoma	*0.776±0.05	*0.782±0.01	0.860±0.01

Table 4. MCC of C4.5, SVM, and BITT algorithm

Data set	MCC		
	C4.5	SVM	BITT
Breast (T1 vs. T2)	*0.564±0.14	*0.108±0.06	0.788±0.06
Breast (T1 vs. T3)	*0.656±0.09	*0.724±0.03	0.919±0.03
Breast (T2 vs. T3)	*0.874±0.06	*0.922±0.01	0.946±0.01
Colon	0.560±0.07	*0.505±0.05	0.675±0.03
Leukemia	0.858±0.05	*0.840±0.04	0.900±0.03
Lung	0.767±0.02	0.904±0.01	0.806±0.02
Lymphoma	*0.432±0.11	*0.207±0.08	0.654±0.04

performance testing process was independently repeated 5 times, which was to verify that the BITT algorithm has a significantly better classifier than the benchmark classifiers, C4.5 and the SVM, by performing a paired Wilcoxon signed-rank test for the right-sided hypothesis testing with $\alpha = 0.05$. The benchmark values of the BITT algorithm, C4.5, and the SVM are implemented using the R programming language (R Core Team, 2015). C4.5 can be used via the J48 package (Hornik *et al.*, 2009; Witten *et al.*, 2016) and the SVM via the e1071 package (Meyer *et al.*, 2017) using the default radial basis kernel function with. We use a fixed value of the SVM regularization parameter gamma equal to 1 over the number of genes.

Results and Discussion

This section describes the experiment and the experimental results from the BITT algorithm. Table 3 shows the average accuracy measurements and the standard deviation of 5 gene expression data sets. The first column is the data set name and next 3 columns show the average accuracy and standard deviation (Mean±SD) of C4.5, the SVM, and the BITT algorithm, respectively. The best value for each data set is shown in bold and an asterisk when the BITT algorithm shows a statistically significant improvement from a 1-tailed (right-sided) Wilcoxon signed-rank test, $p < 0.05$. These results reveal that the average accuracy of the BITT algorithm achieves the highest

value in the colon, leukemia, and lymphoma data sets, but not for the lung cancer data set. Since the breast cancer data set consists of 3 classes, the one against one experiment is applied. The average accuracy of the BITT algorithm for all experiments of the breast cancer data set have significantly improved compared with C4.5 and the SVM. There is only the lung cancer data set for which the SVM classifier shows the highest average accuracy.

The MCC values of the classifiers are shown in Table 4. The BITT algorithm outperforms for the colon, leukemia, lymphoma, and breast cancer in all experiments of one against one. There is only the lung cancer data set for which the SVM is shown as the highest value in all classifiers. The Wilcoxon signed-rank test reveals a significant improvement in most of the comparisons of the BITT algorithm with the other classifiers. There are only the colon and leukemia comparisons with C4.5 for

which the BITT algorithm is not a statistically significant improvement.

Figures 7 and 8 show the box-plots comparison of the accuracy and MCC measures, respectively, of the classifiers in the gene expression data sets which were generated from the 5 repeatedly and independently selected, trained, and tested data in the 5-fold cross validation tests. The outliers are shown with the asterisks. There is only the lung cancer data set for which the BITT algorithm shows lesser average values in the accuracy and MCC measures than the SVM. This may be caused by the highest numbers of genes in the gene expression data set of the lung cancer. The direction of the core vector in the BITT algorithm was disturbed from the less-significant gene in the dataset. As in the breast cancer dataset, the genes which originated from the 24,881 genes (van 't Veer *et al.*, 2002) and which were modified using the feature selection technique to reduce the selected

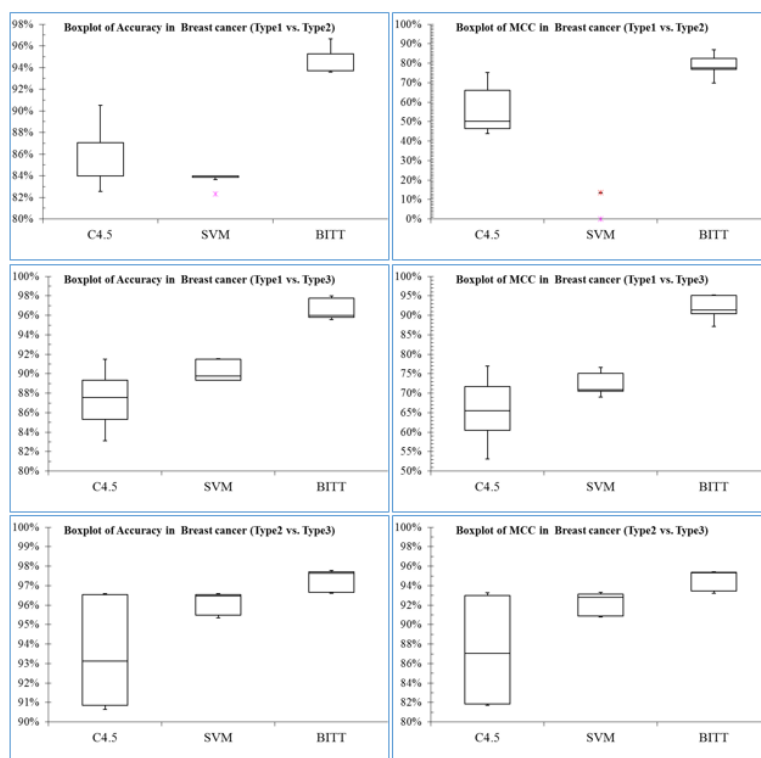


Figure 1. The box-plots of accuracy and MCC for the breast cancer data set

genes (Hoshida *et al.*, 2007) showed the highest accuracy and MCC values in all the experiments.

The MCC measure selected in this study is used to reveal the performance of the classifiers when the data set did not define the actual positive class. The BITT algorithm shows good results in all the data sets except in the lung cancer data set, but especially in the lymphoma data set. In the lymphoma data set, the different value between the SVM and the BITT algorithm is only 7.8% for the average of the accuracy measure but it is 44.73% for the average of the MCC measure.

Conclusions

This paper proposes the BITT algorithm, which is an extension of the RBTP algorithm, that was successful in classifying most cancer types. It is a tree-like structure which is composed of nodes and branches. Each node contains multiple tubes for each class while the branches are identified from the intersection of the multitubes. The recursive partitioning algorithm is used to construct the BITT algorithm node by node from the root to the leaves. In the experiments, the BITT algorithm showed a highly accurate performance of

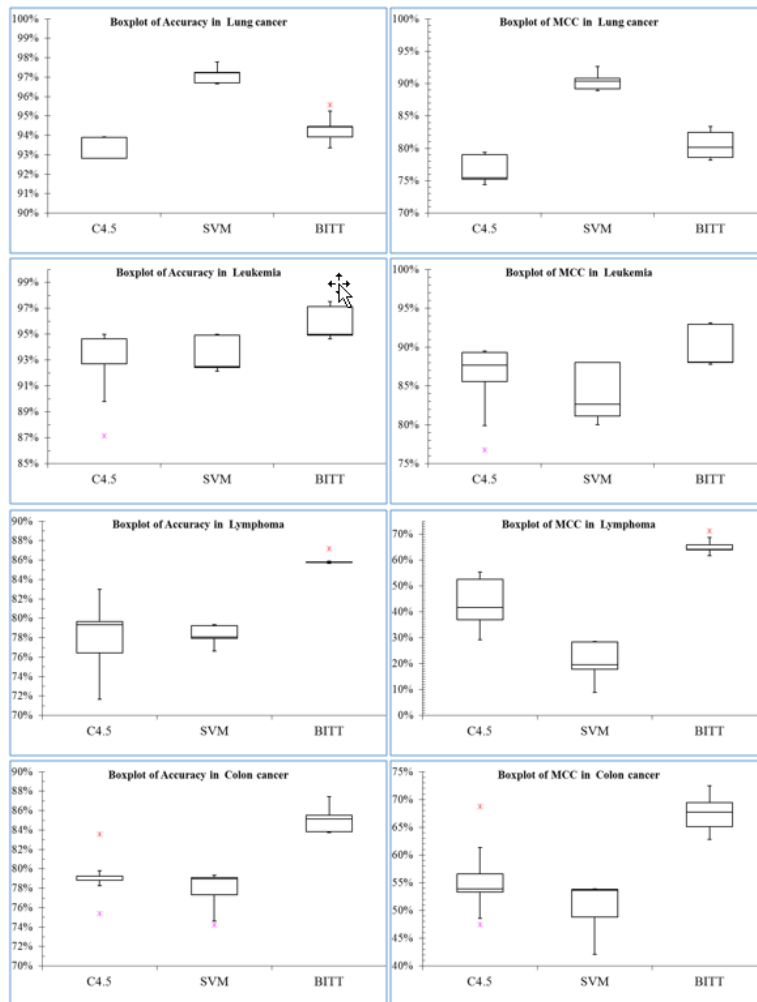


Figure 8. The box-plots of accuracy and MCC for the lung, leukemia, lymphoma, and colon data sets

various cancer data sets, except for the lung cancer which contains a large number of genes (columns) which affect the accuracy of the BITT algorithm. The Wilcoxon signed-rank test was performed on each case to confirm the statistical significance of those experiments. To remedy this situation, it may be necessary to perform anomaly detection or cluster analysis to segment samples into appropriate groups of samples (Ben-Dor *et al.*, 2003; Painsky and Rosset, 2014). This work did not solve the curse of dimensionality problems of the BITT algorithm to find the appropriate interval of the number of genes selected in the experiments. In the future, the preprocessing steps, such as feature selection and transformation, should be applied before using the BITT algorithm to reduce the dimensionality problem.

Acknowledgements

The authors acknowledge the Applied Mathematics and Computational Science curriculum, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University for full support and facilities during the work.

References

- Ben-Dor, A., Chor, B., Karp, R., and Yakhini, Z. (2003). Discovering local structure in gene expression data: the order-preserving submatrix problem. *J. Comput. Biol.*, 10:373-384.
- Bennet, J., Ganaprakasam, C., and Kumar, N. (2015). A hybrid approach for gene selection and classification using support vector machine. *Int. Arab J. Inf. Technol.*, 12, 695-700.
- Cho, S.-B. and Won, H.-H. (2003). Machine learning in DNA microarray analysis for cancer classification. *Proceedings of the First Asia-Pacific Bioinformatics Conference on Bioinformatics 2003*, 19:189-198.
- Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M., and Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16:906.
- Hong, J.-H. and Cho, B.-S. (2006). Multi-class cancer classification with OVR-support vector machines selected by naïve Bayes classifier. *Lect. Notes Comput. Sci.*, 2:155-164.
- Hornik, K., Buchta, C., and Zeileis, A. (2009). Open-source machine learning: R meets Weka. *Computation. Stat.*, 24:225-232.
- Hoshida, Y., Brunet, J.-P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2007). Subclass mapping: Identifying common subtypes in independent disease data sets. *PloS One*, 2:e1195.
- Kanchanasuk, S. and Sinapiromsaran, K. (2016). Recursive binary tube partitioning for classification. *Proc. Adapt. Learn. Opt.*, 5:99-107.
- Lewis, D.P., Jebara, T., and Noble, W.S. (2006). Support vector machine learning from heterogeneous data: an empirical analysis using protein sequence and structure. *Bioinformatics*, 22:2,753-2,760.
- Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., and Yger, F. (2018). A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update. *J. Neural Eng.*, 15:031005.
- Matthews, B.W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *BBA-Protein Structure*, 405:442-451.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.C., and Lin, C.C. (2017). e1071: Misc Functions of the Department of Statistics, Probability Theory Group. R Foundation for Statistical Computing, Vienna, Austria.
- Mohamad, M.S. and Deris, S. (2005). A hybrid of genetic algorithm and support vector machine for features selection and classification of gene expression microarray. *Int. J. Comput. Intel. Appl.*, 5:91-106.
- Mukherjee, S. (2003). Classifying microarray data using support vector machines. In: *A Practical Approach to Microarray Data Analysis*. Berrar, D.P., Dubitzky, W., and Granzow, M. (eds). Springer Science+Business Media, Boston, MA, USA, p. 166-185.
- Nguyen, D.V. and Rocke, D.M. (2002). Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics*, 18:39.
- Painsky, A. and Rosset, S. (2014). Optimal set cover formulation for exclusive row biclustering of gene expression. *J. Comput. Sci. Technol.*, 29:423-435.
- Quinlan, J.R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, Burlington, MA, USA, 302p.
- R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria., 55:275-286.
- Shipp, M.A., Ross, K.N., Tamayo, P., Weng, A.P., Kutok, J.L., Aguiar, R.C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G.S., Ray, T.S., Koval, M.A., Last, K.W., Norton, A., Lister, T.A., Mesirov, J., Neuberg, D.S., Lander, E.S., Aster, J.C., and Golub, T.R. (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.*, 8:68-74.
- Tarca, A.L., Romero, R., and Draghici, S. (2006). Analysis of microarray experiments of gene expression profiling. *Am. J. Obstet. Gynecol.*, 195:373-388.

- van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., and Friend S.H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415:530-536.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bull.*, 1:80-83.
- Witten, I.H., Frank, E., Hall, M.A., and Pal, C.J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. 4th ed. Morgan Kaufmann Publishers, Burlington, MA, USA, 654p.
- Wu, X. and Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. CRC Press, Boca Raton, FL, USA, 232p.
- Zhao, Y-H., Wang, G-R., Yin, Y., and Xu, G-Y. (2007). A novel approach to revealing positive and negative co-regulated genes. *J. Comput. Sci. Technol.*, 22:261-272.