Global Journal of Computer Science and Technology
Hardware & Computation
Volume 13 Issue 1 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Modification of Support Vector Machine for Microarray Data Analysis

By Vrushali Dipak Fangal & Dr. Sk. Sarif Hassan
Indian School of Mines, India

Abstract - The role of protuberant data analysis in selection of certain genes having distinctive level of activities between conditions of interest i.e diseased gene and normal genes is very significant. Now-a-days it is become a standard in gene analysis that microarray of DNA is a crucial data preparation step in systemization and other biological analysis. We consider the problem of constructing an accurate prediction rule for separating the different labels of genes in microarray gene expression data. Use of SVM in such data analysis is not new but it is not up to the mark we desire. So in this manuscript, we have tried to modify Support Vector Machine (SVM) for better accuracy in cancer genes systemization. Here we have modified SVM to account for gene redundancy and keep a check on it. In the other approach, instead of keeping bias a constant in SVM, we have tried to modify SVM by bias variation which we call as Orthogonal Vertical Permutator (OVP).

Keywords : support vector machine, microarray, redundancy, orthogonal vertical permutator.

GJCST-A Classification : C.1.2

MODIFICATION OF SUPPORT VECTOR MACHINE FOR MICROARRAY DATA ANALYSIS

Modification of Support Vector Machine for Microarray Data Analysis

Vrushali Dipak Fangal [α] & Dr. Sk. Sarif Hassan [σ]

Abstract - The role of protuberant data analysis in selection of certain genes having distinctive level of activities between conditions of interest i.e diseased gene and normal genes is very significant. Now-a-days it is become a standard in gene analysis that microarray of DNA is a crucial data preparation step in systemization and other biological analysis. We consider the problem of constructing an accurate prediction rule for separating the different labels of genes in microarray gene expression data. Use of SVM in such data analysis is not new but it is not up to the mark we desire. So in this manuscript, we have tried to modify Support Vector Machine (SVM) for better accuracy in cancer genes systemization. Here we have modified SVM to account for gene redundancy and keep a check on it. In the other approach, instead of keeping bias a constant in SVM, we have tried to modify SVM by bias variation which we call as Orthogonal Vertical Permutator (OVP).

Keywords : support vector machine, microarray, redundancy, orthogonal vertical permutator.

I. Introduction

The theory of support vector machines (SVMs), which is based on the idea of structural risk minimization (SRM), is a new classification technique and has drawn much attention on this topic in recent years (Burges, 1998; Cortes and Vapnik 1995; Vapnik, 1995, 1998). The good generalization ability of SVMs is achieved by finding a large margin between two classes (Bartlett and Shawe-Taylor, 1998; Shawe-Taylor and Bartlett, 1998). In many applications, the theory of SVMs has been shown to provide higher performance than traditional learning machines (Burges, 1998) and has been introduced as powerful tools for solving classification problems. Since the optimal hyper-plane obtained by the SVM depends on only a small part of the data points, it may become sensitive to noises or outliers in the training set (Boser et al., 1992; Zhang, 1999).

To solve this problem, one approach is to do some preprocessing on training data to remove noises or outliers, and then use the remaining set learn the decision function (Cao et al., 2003).This method is hard to implement if we do not have enough knowledge about noises or outliers. In many real world applications, we are given a set of training data without knowledge about noises or outliers. There are some risks to remove the meaningful data points as noises or outliers.

Support Vector Machines have gained much attention in recent years due to their better predictability and ability to theoretically project any data to infinite dimension. It works on the simple basis of separating classes using a hyper plane.

In present scenario, any classification method has to deal with thousands of genes provided by micro array data. This is real test for and classification method. Neural networks have shown high potential in dealing with huge amount of data but it cannot overcome redundancy problem. Many highly correlated genes play similar role in classification while many of them could be omitted. Support Vector Machines also could not account for this problem in current form. In SVM, the weights of any two highly correlated features will be quite near and thus both can play significant role in classification. It is a major hindrance in feature selection.

In this paper, we present two different approaches for improvement of classification accuracy for linear SVMs. In the first method, redundancy control has been targeted for improve the classification rate. For checking a control on redundancy an matrix 'A' has been introduced in the optimization problem. This matrix keeps a check on weight of a feature according to its correlation with other features. It will be discussed in details in later section of paper.

The second method is also an approach to improve the classification performance of linear SVM. It is based on adjustment of bias value in SVM. The results have encouraged us for further probe.

In the paper, Section 2 describes the architecture of normal Support vector machine. Section 3 compares the architecture of normal SVM and modified SVM for controlling the redundancy. Section 4 describes the other method for improving the classification accuracy called Orthogonal Vertical Permutator. Section 5 and 6 discusses experimentation and the results obtained respectively.

II. SVM and Proposed Modification in SVM

a) SVM

Support Vector Machines (SVM's) are learning methods used for binary classification of data. The basic idea is to find a hyper-plane separating n-dimensional

Author α : Department of Computer Science and Engineering, Indian School of Mines, Dhanbad, Jharkhand-826004, India.
E-mail : vrush2elu@ismu.ac.in
Author σ : Institute of Mathematics & Applications Bhubaneswar, India.
E-mail : sarimif@gmail.com

data perfectly into two classes. Since the example data is often not linearly separable, SVM introduce a "kernel induced feature space" which casts data into a higher dimensional space where data is separable. SVM plays a major role in eliminating computational complexity and over fitting (Crammer, Koby, 2001, Drucker, Harris, 1996, Ferris, Michael C, 2002 and T. S. Furey et al. 2000).

We are given l training samples {xi, yi}, i = 1,…., l , where each sample has d inputs $(xi \in R^d)$, and a class label with one of two values (yi∈{-1,1}).Now, all hyper- planes in R^d are parameterized by a vector (w),and a constant (b), expressed in the equation $w.x+b=0$ where w is orthogonal to the hyper-plane.

Given such a hyper-plane (w, b) that separates the data, this gives the function

$$f(x) = sign(w.x+b)$$

which correctly systemizes the training. However, a given hyper-plane represented by (w, b) is equally expressed by all pairs $\{\lambda w, \lambda b\}$ for $\lambda \in R+$. So we define the canonical hyper-plane to be that which separates the data from the hyper-plane by a distance of at least 1. That is, we consider those that satisfy:

xi .w + b > +1whenyi = +1 and $xi.w + b < -1$

when or more compactly: yi(xi .w + b) >1 $\forall i$.

We can frame this as an optimization problem as:

Minimize in (w,b): $||w||$ subject to $(for\ any\ i{=}1,\ldots,n) yi(w.xi-b) \geq 1$

b) Modified SVM

Before we start the modification over the existing SVM let us understand the method of generating the matrix A.

i. *Generation of 'A' Matrix*

1. On basis of property of features like correlation or mutual information.
2. Using a function of importance or unique data points.
3. We may also use something like gradient descent method.

If $w_i = \frac{w_i}{1+\beta_i}$;

Then $E = (Calculated\ Output - Actual\ Output)^2$;

So, $E = (w^T x_i + b - O_i)^2$

Then, $\frac{\partial E}{\partial \beta} = \sum_{i=1} 2[w^T x_i + b - O_i][\begin{matrix} a1 \\ a2 \\ an \end{matrix}]$

$$a_d = \frac{x_{id}.(-w_d)}{(1+\beta_d)^2}$$

Consider the optimization problem in SVM. We introduce a matrix 'A' of order nxn where 'n' is number of features. If we minimize this optimization problem, the weight vector obtained is different from normal SVM weight vector and we can keep a check on redundancy depending on 'A' matrix.

$$m\ in\ \frac{1}{2}||w||^2 + C\sum_{i-1}^{l}\xi_i + \frac{1}{2}w^T A w$$

$$s.t. y_i(w^T x_i + b) \geq 1 - \xi_i \text{where} A = \begin{pmatrix} a_{11} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & a_{mn} \end{pmatrix} \text{is a diagonal matrix.}$$

Introducing Lagrange's multiplier and converting to dual form

$$\phi(w,b,\xi,\alpha,\beta) = \frac{1}{2}||w||^2 + C\sum_{i-1}^{l}\xi_i + \frac{1}{2}w^T A w - \sum_{i=1}^{l}\alpha_i[y_i(w^T x_i + b) - 1 + \xi_i] + \sum_{i=1}^{l}\beta_i\xi_i$$

$$\frac{\partial\phi}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l}\alpha_i y_i = 0$$

$$\frac{\partial\phi}{\partial\xi} = 0 \Rightarrow \alpha_i + \beta_i = C$$

$$\frac{\partial\phi}{\partial\xi} = 0 \Rightarrow w = B^{-1}\sum\alpha_i y_i x_i = \frac{\sum\alpha_i y_i x_i}{1+\alpha_i}$$

Change in Hessian Matrix is-

$$H = y_i y_j x_i x_j$$

$$H = y_i y_j \frac{x_i}{1+a_i} x_j$$

ii. *Comparing Architecture of SVM with Modified SVM*

The layout of normal SVM has been shown below. A separating hyper-plane in canonical form must satisy the following constraints,

$$yi[\langle w,xi\rangle + b] \geq 1,\ i=1,\ldots,l.$$

The distance d(w, b; x) of a point x from the hyper-plane (w, b) is

$$d(w,b;x) = \frac{|\langle w,x^i\rangle + b|}{||w||}$$

$$p(w,b) = \min_{x^i;y^i=-1} d(w,b;x^i) + \min_{x^i;y^i=1} d(w,b;x^i)$$

$$p(w,b) = \min_{x^i;y^i=-1}\frac{|\langle w,x^i\rangle + b|}{||w||} + \min_{x^i;y^i=1}\frac{|\langle w,x^i\rangle + b|}{||w||}$$

$$p(w,b) = \frac{1}{||w||}\left(\min_{x^i;y^i=-1}|\langle w,x^i\rangle + b| + \min_{x^i;y^i=1}|\langle w,x^i\rangle + b|\right)$$

$$p(w,b) = \frac{2}{||w||}$$

Global Journal of Computer Science and Technology (A) Volume XIII Issue I Version I Year 2013

Hence, the hyper-plane that optimally separates the data is the one that minimizes

$$\phi(w) = \frac{1}{2}||w||^2$$

This is solved by using Lagrange's multipliers.

$$\phi(w, b; \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{l} \alpha_i(y_i[< w, x^i > +b] - 1),$$

Where α are the Lagrange multipliers.The Lagrangian has to be minimised with respect to w, b and minimised with respect to $\alpha \geq 0$. Classical Lagrangian duality enables the primal problem,

$$\max_{\alpha} W(\alpha) = \max_{\alpha}(\min \phi(w, b; \alpha))$$

The minimum with respect to w and b of the Lagrangian, ϕ , is given by,

$$\frac{\partial\phi}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0$$

$$\frac{\partial\phi}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{l} \alpha_i y_i x_i$$

Hence, the dual problem is

$$\max_{\alpha} W(\alpha) = \max_{\alpha} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j < x_i, x_j > + \sum_{k=1}^{l} \alpha_k$$

And hence the solution to the problem is given by

$$\alpha^* = arg \min_{\alpha} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j < x_i, x_j > - \sum_{k=1}^{l} \alpha_k$$

With constraints,

$$\alpha_i \geq 0 \quad i = 1, \ldots\ldots, l$$

$$\sum_{j=1}^{l} \alpha_j y_j = 0$$

This equation is can be represented as a quadratic form.

iii. *Orthogonal Vertical Permutator*

Orthogonal Vertical Permutator is a reformation of SVM. In OVP, we vary the bias value of SVM which results in vertical permutations of the hyper-plane resulting from SVM. This section of paper focuses on the bias value 'b' in SVM framework. Its theoretical inspiration is being discussed in following section. Bias is the constant term which is added in decision making equation.

Figure 1 : Depicting the concept of OVP in SVM

a. *Adjustment of Bias Value in SVM*

Consider two concentric circles with each circle representing same class as in Fig. 1. Here, we compare the correct output rate of the output generated by SVM and OVP.

Let the radius of inner circle be 'r' and radius of outer circle be n times 'r' i.e 'nr'. Consider each point on the circle represents a sample.

b. *Correct Classification by SVM*

Correct classification$=\pi r + \pi n r$

$A = (n + 1)\pi r$ *which is half circumference of each circle*

c. *Correct Classification by OVP*

Inner circle is classified correctly.

Correct classification$=2\pi r +$ *Correct classification of outer circle*

We need to find θ to find the correct classification of outer circle. $\theta = Arc\ Radius$

$$\theta = \frac{Arc}{Radius}$$

$$\frac{\pi}{2} - \frac{\theta}{2} = sin^{-1}(\frac{r}{nr})$$

$$\frac{\pi}{2} - \frac{\theta}{2} = sin^{-1}(\frac{1}{n})$$

Therefore,

$$Arc = \theta \times Radius$$

$$Arc = \left(\pi - 2sin^{-1}\frac{1}{n}\right) \times nr$$

$$Arc = \left(\pi nr - 2nrsin^{-1}\frac{1}{n}\right)$$

Thus, the total correct output of by OVP is-

$$B = 2\pi r + n\pi\pi r - 2nrsin^{-1}\frac{1}{n}$$

Difference in the correct output by SVM and OVP is

$$B - A = \left(2\pi r + n\pi\pi r - 2nr\sin^{-1}\frac{1}{n}\right) - (n+1)\pi r$$

$$B - A = \left(\pi r - 2nr\sin^{-1}\frac{1}{n}\right) \geq 0$$

Since, SVM generates the hyper-plane with the best possible slope, here we have adjusted the bias value to shift this plane using minimization of classification error of both classes. Therefore it can be seen that classification error is less in the later case as compared to the normal SVM. For realizing this plane, an approach similar to gradient descent is used. Bias value is changed by a fraction of its current value depending on the minimization of error.

The error in this case is defined in a different way than in usual case.

Normally error is defined as,

$$Error = \frac{Total\ Misclassification\ of\ Class(-1)\ and\ Class(1)}{Total\ Number\ of\ Samples}$$

But in this case, we have defined Error as

$$Error = \frac{Total\ Misclassification\ of\ Class(-1)}{Total\ Number\ of\ Samples\ of\ Class(-1)} + \frac{Total\ Misclassification\ of\ Class(+1)}{Total\ Number\ of\ Samples\ of\ Class(+1)}$$

Both error rates are quite different. In first case the, each error has absolute importance and is equally important. But in the other case, the error rate for each class is different and its importance is related to the number of samples in its class. In second case, one class may be classified to very high accuracy at the expense of the other. It leads to higher probability of accuracy rate for one class.

SVM is used to categorize datasets into binary data. All the hyper-planes separating the data into two groups are orthogonal to vector w. The variation of bias gives rise to various permutations of the hyper-planes along the vertical. Our model gives rise to vertical permutations of the orthogonal hyper-planes and hence the Orthogonal Vertical Permutator is named.

III. Result and Analysis

This section gives the comparison of percentage accuracy of SVM with modified SVM in Table-I against the sigma values of A matrix. It can be inferred that the modification offers a better accuracy over SVM. The second table depicts a comparison of percentage of accuracy of SVM and OVP-SVM. The percentage accuracy with shifted bias value is better than normal bias value. Figure 2 and figure 3 gives the graphical representation of application of SVM and OVP-SVM on concentric circle dataset and spiral dataset respectively.

A Matrix	Modified SVM	SVM
All Sigmas	75.85	75.25
Sigma>0.75	75.05	74.45
Selected Sigmas	76.3	75.4

Table 1 : Comparing Accuracy Results of SVM and Modified SVM

Dataset	SVM	OVP-SVM
Concentric Circle	49.7%	63%
Spiral	49.8%	53.05%

Table 2 : Comparing Accuracy Results of SVM and OVP-SVM

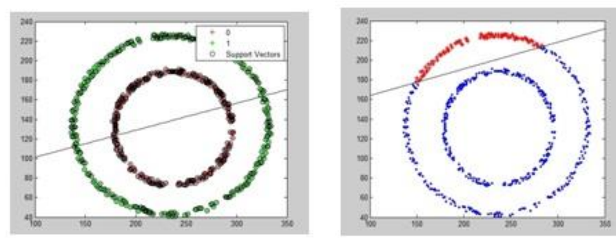

Figure 2 : Results of SVM and OVP-SVM on Concentric Circle Dataset

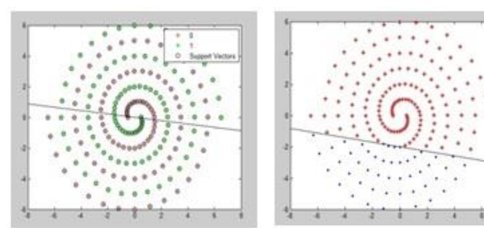

Figure 3 : Results of SVM and OVP-SVM on Spiral Dataset

The comparison analysis of both classification methods is done on the benchmark datasets. Each dataset is validated using double cross fold approach. Linear SVM is used for classification. Therefore only one parameter needs to be tuned i.e. the 'C' which accounts for soft margin classification. If training data was not provided separately then data was analyzed using 5 fold double cross validation. The data was divided into four parts of training data and one part of testing data. This training data was again five folded with four folds for actual training and one fold for parameter adjustment. All the datasets are available at UCI Machine Learning repository [6]. Table 1 shows the effect of change in matrix 'A' on sonar dataset. In rest of cases results are obtained by creating matrix 'A' is made by summing the correlation coefficient of a feature with rest of the features.

The experimentation of SVM with changed value of bias is performed on two datasets. The result of concentric circle dataset and of Spiral Dataset is shown in table 2.

The results obtained in both methods outperform the normal SVM and have different advantages. Modified SVM is immune to redundancy and OVP helps in improvising the classification accuracy of SVM and can be beneficial in multi class datasets. The processing was done on Matlab R2009.

References Références Referencias

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp. 68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. Crammer, Koby; and Singer, Yoram (2001). "On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines". J. of Machine Learning Research 2: 265–292. R. Nicole.
5. Drucker, Harris; Burges, Christopher J. C.; Kaufman, Linda; Smola, Alexander J.; and Vapnik, Vladimir N. (1997); "Support Vector Regression Machines", in Advances in Neural Information Processing Systems 9, NIPS 1996, 155–161, MIT Press.
6. Ferris, Michael C.; and Munson, Todd S. (2002). "Interior-point methods for massive support vector machines". SIAM Journal on Optimization13 (3): 783–804.
7. T. S. Furey et al. (2000), support vector machine classfication and validation of cancer tisue samples using microarray expressino data. Bioinformatics, 16 (10): 906-914.
8. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
9. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

This page is intentionally left blank