Available online at www.sciencedirect.com

ScienceDirect

Procedia Technology 23 (2016) 369 – 375

Procedia Technology

3rd International Conference on Innovations in Automation and Mechatronics Engineering, ICIAME 2016

Support Vector Machine – a Large Margin Classifier to Diagnose Skin Illnesses

Krupal S. Parikh[a], Trupti P. Shah[b]

[a,b]*Department of Applied Mathematics, Faculty of Technology & Engineering,*
The M. S. University of Baroda, Vadodara, Gujarat State, India - 390 001

Abstract

Support Vector Machine (SVM) have been very popular as a large margin classifier due its robust mathematical theory. It has many practical applications in a number of fields such as in bioinformatics, in medical science for diagnosis of diseases, in various engineering applications for prediction of model, in finance for forecasting etc. It is widely used in medical science because of its powerful learning ability in classification. It can classify highly nonlinear data using kernel function. This paper proposes and analyses diagnostic model to classify the most common skin illnesses and also provide a useful insight into the SVM algorithm. In rural areas where people are generally treated by paramedical staff, skin patients are not subject to proper diagnosis resulting in mistreatment. We think SVM is a good tool for proper diagnosis. This paper uses various kernels for classification and achieving the best accuracy of 95.39 % .

Peer-review under responsibility of the organizing committee of ICIAME 2016

Keywords: Support Vector Machine; Accuracy; F-Score; G-Score

1. Introduction

Support Vector Machine (SVM) is kernel-based supervised learning algorithm, which is the combination of Machine learning theory, optimization algorithms from operation research and kernel techniques from Mathematical analysis. A good generalization of a classifier is achieved when it minimizes training error along with higher testing accuracy for unknown testing dataset. The training algorithm of SVM maximizes the margin between the training data and class boundary, removing some meaning less data from the training dataset. So, the resulting decision function depends only on the training data called support vectors, which are closest to the decision boundary. Thus SVM maximizing the boundary by minimizing the maximum loss and giving good accuracy compared to the

Peer-review under responsibility of the organizing committee of ICIAME 2016
doi:10.1016/j.protcy.2016.03.039

classifier which are based on the minimizing the mean squared error [1]. It is also effective in high dimensional space where number of dimension is greater than the number of training data. SVM can separate the classes which cannot be separated by linear classifier. SVM is kernel based method. It uses the kernel induced feature space [2]. Using a kernel function it transforms data from input space into a high-dimensional feature space in which it searches for a separating hyper plane. So, that nonlinear data can also be separated using hyper plan in high dimensional space. This takes a lot of computation power. But SVM overcome this problem using kernel trick. In SVM kernel functions are defined in reproducing kernel Hilbert space (RKHS)[3]. Hilbert space is complete inner product space so similarity between training data points are measured by inner product which is less expensive computationally. Also, kernels are Mercer's kernel [4], i.e., positive semi definite kernel and due to the Mercer's kernel SVM gives global optimum.

High learning ability, good generalization in classification and regression makes SVM most popular learning algorithm in many real-life applications such as bioinformatics, electrical load forecasting[5], pattern recognition, image processing, field of hydrology[6]. SVM is used to predict mechanical property such as hot-rolled plain carbon steel[7], to build credit scoring models assessing the risk of default of clients [8],in fault diagnosis[9], for forecasting failures and reliability in engine system[10]. It is also used to evaluate level of coal mine underground environment[11], in classification of drug and nondrug problem [12] , to diagnosis diabetes and erythematous disease [13,15], in drug design, in qualitative and quantitative prediction from sensor data etc.[14].

Skin diseases such as Bacterial Infection, Fungal Infection, Eczema and Scabies are common problems particularly in underdeveloped countries. Large population commonly suffers from these diseases. Such skin diseases are commonly encountered by medical and paramedical staff at primary health centers, community health centers, referral hospitals as well as in specialized hospitals. There is a definite need for proper diagnosis and treatment for such disorders. Because of the improper diagnosis many times they are treated incorrectly and by mixture of antibacterial, antifungal and steroid preparation locally. Such treatment is hazardous to the society. Due to wrong diagnosis, improper treatment makes the disease more complicated and later on it is difficult for dermatologist to give proper treatment. We believe that at a primary stage computer assisted diagnosis is necessary to avoid major complication at the later stage.

In this study, we have used Support Vector Machines to diagnosis these diseases. SVM is originally designed for Binary Classification. We have used one-to-one algorithm for our multiclass data. The database was obtained from Department of Skin & V.D., Shrikrishna Hospital, Karamsad, Gujarat, India.

2. Support Vector Machine

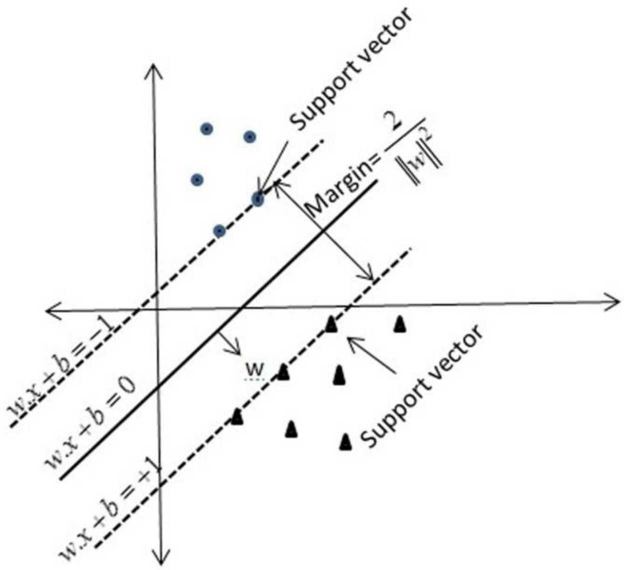
SVMs are among the best "off-the-shelf" supervised learning algorithms[16]. It is kernel based supervised learning algorithm for binary classification problem. It separates the two classes using kernel function which is induced from the training data set. The goal is to produce a classifier that will work well on unseen examples, i.e. give good generalization.
Let there be m training examples $(x_i, y_i), y_i = \pm 1, \ i = 1,2,3,\ldots m$.

Then there exist a hyper plane $\mathbf{w}.\,\mathbf{x} + b = 0$, which separate the positive and negative training examples using the decision function :

$$f(x) = sign(\mathbf{w}.\,\mathbf{x} + b), \text{where } \mathrm{sign}(\mathrm{x}) :- \begin{cases} -1, & \text{if } \ \mathrm{x} < 0 \\ 0, & \text{if } \ \mathrm{x} = 0 \\ 1, & \text{if } \ \mathrm{x} > 0 \end{cases} \tag{1}$$

where, $\mathbf{w}$ is the normal to the hyper plane which is known as weight vector and b is called the bias. We see that $y_i(\mathbf{w_i x_i} + b) > 0, \ \forall \ i = 1,2,3,\ldots m$.

Figure 1 Maximum-margin hyper plane. Training data (instance) on the margin are called the support vectors.



Implicitly define $(\mathbf{w},b)$ such that $(\mathbf{w.x}+b)=1$ for positive class and $(\mathbf{w.x}+b)=-1$ for negative class (see figure 1), then there be two hyper planes and the region between these hyper planes is called the margin band, given by $\frac{2}{\|\mathbf{w}\|^2}$,which is to be maximize or

Minimize $\frac{1}{2}\|\mathbf{w}\|^2$ subject to the constraints: $y_i(\mathbf{w.x_i}+b)\geq 1,\ \forall\ i=1,2,3,...m.$ (2)

Most real life dataset contains noise. Using a soft margin the effects of outliers and noise can be reduced. By introducing the soft margin with marginal error ξ_i ,the objective function becomes[17],

Minimize $\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{m}\xi_i$ with, $y_i(\mathbf{w.\,x_i}+b)\geq 1-\xi_i,\ \xi_i\geq 0\ \forall\ i=1,2,3,...m.$ (3)

The parameter C controls the tradeoff between the marginal error and testing error.

Karush-Kuhn-Tucker (KKT) conditions are necessary conditions for nonlinear optimal problem.

The primal problem is converted into dual problem and applying KKT conditions we obtain $\mathbf{w}=\sum_{i=1}^{m}\alpha_i y_i \mathbf{x}_i$ (4)

Using kernel function decision function becomes $f(\mathbf{x}) = sign\left(\sum_{i \in Sv} \alpha_i y_i K(\mathbf{x}, \mathbf{x_i}) + b\right)$, with the objective function[18]:

$$\text{Maximize } w(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(\mathbf{x_i}, \mathbf{x_j}), \ \forall \ i = 1,2,3,\ldots m \quad \text{with } \alpha_i \geq 0, \sum \alpha_i y_i = 0. \tag{5}$$

where, $K(x_i, x_j) = \phi(x_i).\phi(x_j)$, $\phi(x)$ is called kernel function, which need not to be known. It is defined by the inner product in feature space. So, the feature space should be inner product space called a Hilbert space. These kernels are Mercer's kernels which are positive semi definite hence global optimum is achieved. The objective function is Quadratic Optimization Problem which can be solved by the Sequential Minimal Optimization (SMO) algorithm.

3. Experiments and Result Analysis

In our study we have used SVM to classify some common skin diseases viz., Bacterial Infection, Fungal Infection, Scabies and Eczema. The data was collected from Department of Skin & V.D., Shrikrishna Hospital, Karamsad, Gujarat, India. We have prepared detailed Proforma under the guidance of leading dermatologist and investigated 470 patients. To find the attributes deep investigation as well as doctor's ideas have been taken care of. The proforma includes 47 features.

There are 47 features and 470 instances. Out of 470 instances 139 instances are for Bacterial Infections, 146 for Fungal Infection, 98 for Eczema and 87 for Scabies. Table 1 show various features which are investigated during our data collection.

Table 1. Input Attributes used for Analysis

	Chief Complaints & OPD				
1.	Pain	2.	Fever	3.	Itching
	Seasonal relation				
4.	Summer	5.	Winter	6.	Monsoon
	Past History				
7.	Diabetes Mellitus	8.	Family History		
	Occupational History				
9.	Hot and humid environment	11.	Excessive sun exposure		
10..	Exposure to irritants				
	Type of Lesion				
12.	Macules	16.	Nodule		
13.	Patches	17.	Plaques		
14.	Papules	18.	Vesicles		
15.	Pustule	19.	Bullae		
	Colour				
20.	Erythematous	22.	Hypopigmented		
21.	Hyperpigmented				

	Associated With				
23.	Lichenification			26.	Scaling
24.	Oozing			27.	Excoriation
25.	Crusting			28.	Discharge
	Shape				
29.	Linear	30.	Annular	31.	Grouped
	Sites				
32.	Webspaces	37.	Abdomen	42.	Back
33.	Wrist	38.	Genitals	43.	Buttocks
34.	Forearm	39.	Thigh	44.	Palms & Soles
35.	Arm	40.	Legs		
36..	Chest	41.	Dorsa of feet		
45.	Hair	46.	Nail	47.	Face

Kernel is the key that determines the performance of the SVM. We have used various kernels to train our data. We have calculated confusion matrix for each kernel functions. Table 2 present general form of the confusion matrix. Tables from 3 to 7 present the confusion matrices for various kernels used. The table would contain the average values for all classes combined.

The SVM parameter (regularization parameter) which controls the tradeoff between the marginal error and testing error is C=4096.

Table 2. Confusion Matrix

	Predicted Positive Class	Predicted Negative Class
Actual Positive Class	True Positive (TP)	False Negative (FN)
Actual Negative Class	False Positive (FP)	True Negative (TN)

Table 3 Confusion Matrix for Linear Kernel

	Predicted Positive Class	Predicted Negative Class
Actual Positive Class	122	19
Actual Negative Class	19	404

Table 4. Confusion Matrix for Polynomial kernel

	Predicted Positive Class	Predicted Negative Class
Actual Positive Class	128	13
Actual Negative Class	13	410

Table 5 Confusion Matrix for Radial Basis Function

	Predicted Positive Class	Predicted Negative Class
Actual Positive Class	128	13
Actual Negative Class	13	410

Table 6 Confusion Matrix for t-Student

	Predicted Positive Class	Predicted Negative Class
Actual Positive Class	127	14
Actual Negative Class	14	409

Table 7 Confusion Matrix for Inverse Multiquadratic

	Predicted Positive Class	Predicted Negative Class
Actual Positive Class	126	15
Actual Negative Class	15	408

Accuracy is an important evaluation of any classifier which is defined as :

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}.$$

For imbalanced data instead of accuracy, F-Score is a useful measurement which is weighted average of the precision and sensitivity. We calculate F-score to measure the performance of the SVM classifier using various kernels and also find G-Score which do not account the size of positive and negative classes and provide a fair comparison.

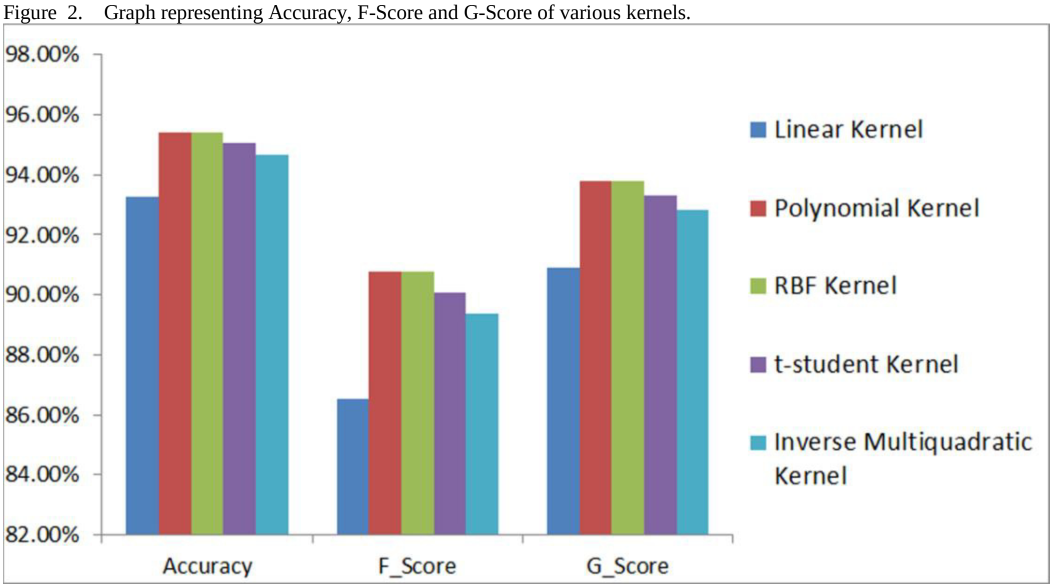
$\text{F-Score} = \frac{2 \text{ x Sensitivity x Specitivity}}{\text{Sensitivity} + \text{Specitivity}}$ and $\text{G-Score} = \sqrt{\text{Sensitivity x Specitivity}}$ where, Sensitivity which also called true positive rate or recall, measure the proportion of positive that are correctly identified and is defined as $\text{Sensitivity} = \frac{TP}{TP + FN}$ and Specificity which is also called the true negative rate measures the proportion of negatives that are correctly identified and is defined as $\text{Specificity} = \frac{TN}{TN + FP}$

Experiments are performed in MATLAB using LIBSVM 3.20[19]. We have randomly selected 70% of our data as training data while rest of 30% data are used for testing and obtained highest accuracy of 95.39 % with F-score 90.78% and G-score 93.80%. Table 8 presents the performance SVM taking various kernels with SVM parameter C=4096.

Table 8 Performance SVM taking various kernels(SVM parameter C=4096)

Kernel Function K(x,y)	Values of kernel Parameters	Accuracy	F-Score	G-Score
Linear: $x^T y + c$	c=10	93.26%	86.52%	90.91%
Polynomial: $(\alpha x^T y + c)^d$	α =2, c=10, d= 3	95.39%	90.78%	93.80%
Radial Basis Function (RBF) $exp\left(-\gamma\|x-y\|^2\right)$	γ =0.1	95.39%	90.78%	93.80%
t-Student: $1/\left(1+\|x-y\|^d\right)$	d=2	95.04%	90.07%	93.32%
Inverse Multiquadratic $1/\left(c^2+\|x-y\|^2\right)$	c=10	94.68%	89.36%	92.84%

Figure 2. Graph representing Accuracy, F-Score and G-Score of various kernels.



4. Conclusion

The paper focuses on the power of kernel based support vector machine in medical diagnosis of some common skin illnesses viz. bacterial infections, fungal infections, eczema and Scabies. The results of our computation recorded in table and plotted graphically indicate that the Radial Basis Function (RBF) kernel and polynomial kernel gives better accuracy than the other kernels for our set of data. It also shows that Kernel methods are very efficient if they are applied in creative ways and can solve a wide range of problems in science and engineering. In future work, more powerful new kernel functions can be proposed to increase the accuracy of the classifier and as a result diagnosis can be done more accurately.

Acknowledgements

Authors are thankful to Shrikrishna Hospital for giving the permission for data collection. Additionally, authors would also like to thank the department of Skin & VD for their help in collecting data.

References

[1] Bernhard Boser, Isabelle Guyon, Vladmir Vapnik. A Training Algorithm for Optimal Margin Classifiers. Proceedings of the fifth annual workshop on Computational learning theory,ACM, 1992, 144-152.

[2] Claudio Carmeli, Ernesto De Vito, Alessandro Toigo. Reproducing Kernel Hilbert Spaces and Mercer theorem. arXiv preprint math/0504071,2008.

[3] Bernhard Scholkopf, and Alexander J. Smola. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press, 2001.

[4] Martin Hofman. Support Vector Machines-Kernel and the Kernel Trick. Houptseminar report 2006.

[5] Wei-Chiang Hong. Electric load Forecasting by support vector machine model. Applied Mathematical Modeling 2009, 33(5), 2444-2454.

[6] Sujay Raghavendra, Paresh Deka. Support vector machine applications in the field of hydrology:A review. Applied soft computing 2014:19,372-386.

[7] Ling Wang, Zhichun Mu, Hui Guo. Applications of support vector machine in the prediction of mechanical property of steel materials. Journal of University of Science and Technology Beijing, Mineral, Metallurgy, Material.2006,13(6):512-515.

[8] Van Gestel, Ir Tony, Bart Baesens, Ir Joao Garcia, Peter Van Dijcke. A support vector machine approach to credit scoring. In forum financier-revue bancaire et financiaire bank en financiewezen- unknown 2003,73-82.

[9] Esmaeil Tafazzoli, Mehrdad Saif. Application of combined support vector machines in process fault diagnosis. American Control Conference 2009.

[10] Xinfeng Zhang, Zhao Yan. Application of support vector machine to reliability analysis of engine systems.TELKOMNIKA Indonesian Journal of Electrical Engineering 2013, 11(7),3552-3560.

[11] Zhou Huaping, Ruixin Zhang. Application of Support Vector Machine Model in Mine Gas Safety Level Prediction.TELKOMNIKA Indonesian Journal of Electrical Engineering 2014, 12(5),4056-4062.

[12] Byvatov Evgeny, Uli Fechner, Jens Sadowski, Gisbert Schneider. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. Journal of Chemical Information and Computer Sciences. 2003,43(6),1882-1889.

[13] Kumari V. Anuja , R. Chitra. Classification of Diabetes Disease Using Support Vector Machine. International Journal of Engineering Research and Applications. 2013. 3(2), 1797-1801.

[14] Ivanciuc Ovidiu. Applications of support vector machines in chemistry. Reviews in computational chemistry. 2007,.23:291.

[15] A. Basu, S. Roy, A. Abraham. A Novel Diagnostic Approach Based on Support Vector Machine with Linear Kernel for Classifying the Erythemato-Squamous Disease. Computing Communication Control and Automation(ICCUBEA) –International conference on 2015,343-347.

[16] Sevakula, Rahul Kumar, and Nishchal K. Verma. Support vector machine for large databases as classifier. Swarm, Evolutionary, and Memetic Computing. Springer Berlin Heidelberg, 2012. 303-313.

[17] Colin Campbell, Yiming Ying, *Learning with Support Vector Machines*, Synthesis Lectures on Artificial Intelligence and Machine Learning #10, Brachman Ronald and Dietterich Thomas,editors,Morgan & Claypool publishers, 2011.

[18] Alexander Smola,Peter Bartlett, Bernhard Scholkopf, Dale Schuurman, *Adavances in Large-Margin Classfiers*,MIT press,1999.

[19] https://www.csie.ntu.edu.tw/~cjlin/libsvm/