International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 1 (2015) pp. 2081-2090

http://www.ripublication.com

An Efficient Feature Reduction Technique For An Improved Heart Disease Diagnosis

[1]G. N. Beena Bethel, [2]Prof. T. V. Rajinikanth, [3]Prof. S. Viswanatha Raju

[1]Assoc. Professor, CSE Dept, GRIET, Hyderabad, India, email_id: beenabethel@gmail.com
[2] Professor, CSE Dept., SNIST, Hyderabad, India. email_id: rajinitv@gmail.com
[3]Professor, CSE Dept., JNTUH (Jagityal), Karimnagar, AP., India. email_id: viswanadha_raju2004@yahoo.co.in

Abstract

Computerized decisions are useful for the physicians for fast and accurate decision making process. Especially for the diseases related to heart, a diagnostic decision plays an important role in saving a human life. An extensive research is however carried out in finding better ways to select the features for the heart disease diagnosis. And then it could be followed by classification so that efficient and fast decisions are possible. We focus on a better feature reduction technique to explore various kernels of the Support Vector Machines classifier and thus improve the accuracy of the decisions to a little extent. Cleveland heart disease dataset taken from UCI repository is used so that comparing with other techniques is done without compromise. Data was divided into training and testing as per the thumb rule which says approximately two thirds as training data and one thirds as test data.

Keywords— Support vector machines, multiclass SVM, Quadratic Programming, Least Squares, Sequential Minimal Optimization, kernel Principal Component Analysis.

Introduction

Heart disease claims to be the number one killer among the other major diseases. The diagnosis of heart disease in itself is a complex and cognitive process and needs to be interpreted more accurately. Although a good number of diagnostic tests are conducted in order to assess whether a patient is suffering from heart disease, most of them are not relevant for the diagnosis. Hence feature selection aids in picking out the features that are essential for the diagnosis. In terms of computing, a huge number of

dimensions lead to curse of dimensionality and hence the number of features needs to be reduced for the efficient diagnosis, which is done using kernel PCA. kPCA is a technique which brings the non-linear data into a linear space and then applies the Principal component analysis to select the features.

The plan of this paper is like: In section II, the related work in the area of heart disease diagnosis, done by various researchers is mentioned. In section III the dataset that was used in this study is being described. In section IV the methodology of how the SVM classifier takes the help of kPCA's feature reduction technique to improve the accuracy of the diagnosis, with all its theoretical aspects were described. In section V the experimental analysis is being placed and section VI concludes the work with summary of the findings.

Related Work

Lot of research has been carried out in the diagnosis of the Cardio Vascular disease. The Cleveland's heart disease data was used in this work and it has thirteen features plus a class label feature. Not all of them contribute to the efficiency of classification due to curse of dimensionality and so the features are to be reduced in order to increase the efficiency of classification. Many such studies could get accuracy below 80 percent but when kernel principal component analysis was used to reduce the number of features it could get accuracy beyond 80 percent.

Sumit Bhatia, Praveen Prakash, and G.N. Pillai in [15] used Integer-Coded Genetic Algorithm to Select Critical Features on Cleveland Data. Yang J. and Honavar, V in [20] used genetic algorithm to select optimal features with back propagation based on multi-criteria optimization in terms of generalization accuracy and the costs associated with features. C.L.Huang, and C.J. Wang, in [8] used a GA-based feature selection and parameters optimization for support vector machines. Bailin Liu, Huiyun Zou, Xi Chen in [2] used rough set theory for reducing the feature set. Yan, H., Zheng, J., Jiang, Y., Peng, C. and Xiao, C., in [19] Selected critical clinical features for heart diseases diagnosis with a real-coded genetic algorithm. Comak, E., Arslan, A. and Ibrahim, T. in [5] have worked on support [1]vector machines for diagnosing the heart valve diseases. Zhao, Chen, Hou, Zheng, & Wang in [21] used backward elimination procedure along with a novel algorithm. Fan & Chaovalitwongse in [7] suggested a novel optimization framework for getting improved feature selection in classification. Soman, Shyam, & Madhavdas, in [14] worked on decision trees and SVM to predict heart disease. K.R. Lakshmi, M. Veera Krishna and S. Prem Kumar in [10] used linear discriminant analysis for feature reduction and have applied various classification techniques for heart disease diagnosis. E. Avci and Ibrahim Turkoglu in [1] worked out an intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases. Mi Hye Song, Jeon Lee, Sung Pil Cho, Kyoung Joung Lee, Sun Kook Yoo have proposed an algorithm for arrhythmia classification using SVM by reducing the

features using Linear Discriminant Analysis which gave better performance than with PCA in [11].

Dataset Description

The dataset that was used for this study is Cleveland heart data from UCI repository which was donated by Detrano [16]. It consists of 76 attributes on the whole but only 14 of them are extensively used by the computational intelligence research. The 14 attributes of the Cleveland dataset along with the values and data types are as follow [9]:

1. Age: age in years (numeric);
2. Sex: male, female (nominal);
3. Chest pain type (CP): (a) typical angina (angina), (b) atypical angina (abnang), (c) non-anginal pain (notang), (d) asymptomatic (asympt) (nominal).

From medical point of view,

(a) Typical angina is the condition in which the past history of the patient shows the usual symptoms and so the possibility of having coronary artery blockages is high [3].

(b) Atypical angina refers to the condition that the patient's symptoms are not detailed and so the probability of blockages is lower [3].

(c) Non-angina pain is the stabbing or knife-like, prolonged, dull, or painful condition that can last for short or long periods of time [6].

(d) Asymptomatic pain shows no symptoms of illness or disease and possibly will not cause or exhibit disease symptoms [12]

4. Trestbps: patient's resting blood pressure in mm Hg at the time of admission to the hospital (numeric);
5. Chol: Serum cholesterol in mg/dl;
6. Fbs: Boolean measure indicating whether fasting blood sugar is greater than 120 mg/dl: (1 = True; 0 = false) (nominal);
7. Restecg: electrocardiographic results during rest. Three types of values normal (norm), abnormal (abn): having ST-T wave abnormality, ventricular hypertrophy (hyp) (nominal);
8. Thalach: maximum heart rate attained (numeric);
9. Exang: Boolean measure indicating whether exercise induced angina has occurred: 1 = yes, 0 = no (nominal);
10. Oldpeak: ST depression brought about by exercise relative to rest (numeric);
11. Slope: the slope of the ST segment for peak exercise. Three types of values upsloping, flat, downsloping (nominal);
12. Ca: number of major vessels (0–3) colored by fluoroscopy (numeric);
13. Thal: the heart status (normal, fixed defect, reversible defect) (nominal);
14. The class attributes: value is either healthy or heart disease (sick type: 1, 2, 3, and 4).

The UCI's heart disease dataset has five class attributes indicating either healthy or one of four sick types. For this paper, multi-class classification is converted into a binary classification, thereby it results in 0 for healthy and 1 for unhealthy cases.

There are 303 cases in the dataset out of which 6 have missing data and are eliminated for our classification and it resulted in 297 cases for our work.

The 297 records which were obtained from UCI repository were normalized such that the data fits into the range of 0 and 1. This is done by using 0-1 normalization which is given by

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

IV. Methodology

We have worked on the data taking all the five distinct classes to be classified as multiclass SVM and also we have classified the data into only two classes, as 0 for 'healthy' cases and 1 for 'unhealthy' cases. Classification was done using various kernels of SVM like linear, polynomial, radial basis function, multi-layer perceptron and quadratic functions. It was implemented for all the 13 features taken together and also for the reduced feature set done through kernel PCA.

4.1 SVM

Support vector machine is a supervised learning algorithm which was first introduced by Vapnik V. N. in 1995[17]. It builds a hyperplane in which it divides all the cases such that the cases of one class are all on one side of the plane and the cases of other class on the other side of the plane. On the whole SVM can classify only two classes.

The essence on which SVM works is to find a hyperplane which maximizes the minimum distance from any of the training data. The optimal margin hyperplane is represented as a linear combination of training points. Consequently the decision function for classification with respect to hyperplane involves only dot products.

The equation of the hyperplane dividing the two classes is given by

$$w^T x = a \tag{2}$$

The equations for the dashed hyperplanes are given by

$$w^T x = a + b \tag{3}$$

$$w^T x = a - b \tag{4}$$

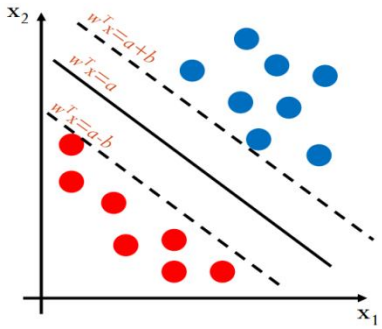

Fig. 1 Hyperplane dividing data into two distinct classes

Various kernels used in our study are:

4.1.1 Linear SVM

Linear Kernel function is given by

$$D = \left\{ (x_i, y_i) \mid x_i \in R^P, y_i \in \{-1,1\} \right\}_{i=1}^{n} \quad (5)$$

Where y_i is either 1 or -1, indicating the class to which x_i belongs. x_i is a P-dimensional vector.

Any hyperplane for the set of points X can be written satisfying the equation,

$$w.x - b = 0 \quad (6)$$

The parameter $\frac{b}{\| w \|}$ determines the offset of the hyperplane from the origin along the normal vector w.

4.1.2 Polynomial Kernel

The most commonly used polynomial kernel function is given by

$$k(x, x') = \left[1 + x^T x' \right]^k \quad (7)$$

Where *k* is the order of the polynomial taken as 3 in our work.

4.1.3 Gaussian Kernel

Another widely used Gaussian radial basis function is given by

$$k(x, x') = \exp\left[-\frac{1}{2} \| x - x' \|^2 \right] / \sigma \quad (8)$$

Where σ is the spread of the Gaussian function.

4.1.4 Multi-layer perceptron kernel

The multi-layer perceptron kernel is built with neural networks where it classifies the n-dimensional input vector by adjusting its weights into one of the two classes indicated by the output nodes of the neural network.

4.1.5 Quadratic kernel

Quadratic kernel used in SVM is similar to that of polynomial with a degree 2 is given by

$$k(x, x') = \left[1 + x^T x' \right]^2 \quad (9)$$

The above specified kernels are used in SVM for predicting the heart disease using sequential minimal optimization, least squares and quadratic programming methods. Each of these methods on various kernels are tested before reducing the features and after reducing the features using kernel Principal Component Analysis. Most researches have used PCA to reduce features but kernel PCA is giving a better accuracy.

4.2 Multiclass SVM

The support vector machines are basically two-class classifiers. Quite often we come across problems involving k>2 classes. Our Cleveland heart disease dataset is one of the same type which has five classes namely, 0 for healthy heart and 1,2,3,4 for the presence of heart disease in that each represents the level of diameter narrowing of the arteries. For classifying more than two classes, we go about combining multiple two-class SVMs so as to construct a multi-class classifier.

There are many approaches to do this multi-class classification and we adopted a commonly used *one-versus-the-rest* approach which was introduced by Vapnik in [18]. We have constructed a multi-class SVM out of K separate SVMs in which k^{th} model $y_k(x)$ is trained using the data from class C_k as positive tuples and the rest of k-1 classes as negative tuples. This method is frequently used for making predictions for new inputs x using

$$y(x) = \max_k y_k(x) \tag{11}$$

4.3 Kernel PCA

Having high dimensional data is prone to curse of dimensionality and so the number of features have been reduced. Most researchers used Principal component analysis for reducing the feature set of the data which is mostly used for linear data. Kernel substitution allows us to express the algorithm in terms of scalar products of the form $x^T x'$ [4] and generalized it by replacing the scalar products with a non-linear kernel. Here we apply kernel substitution to principal component analysis to obtain a non-linear generalization called kernel PCA [13].

Due to the nonlinear transformation, kPCA is able to include non-linearity in the data. Since the non-linear transformation used in kPCA is unknown, it is difficult to transform the Eigen vector from the feature space into the input space. Hence we use kernel trick to approximate the reconstructed data into the data space [17].

Experimental Results

The data obtained from the UCI repository was first normalized and then classified into two classes where 0 is for num = 0 (i.e., healthy) and 1 for num = 1, 2, 3, 4 (i.e., unhealthy) cases.

Table I Two class SVM classification without feature reduction

SVM without PCA or kPCA	Least Squares	Sequential minimal optimization	Quadratic programming
Linear	81.4433	83.5052	80.4124
Polynomial	75.2577	73.1959	73.1959
RBF	69.0722	70.1031	69.0722
Quadratic	71.1340	71.1340	-
Multilayer perceptron	52.5773	81.4433	-

Table II Two class SVM classification using PCA

SVM using PCA	Least Squares	Sequential minimal optimization	Quadratic programming
Linear	81.4433	82.4742	81.4433
Polynomial	80.4124	77.3196	77.3196
RBF	82.4742	85.5670	82.4742
Quadratic	79.3814	81.4433	-
Multilayer perceptron	55.6701	83.5052	-

Table III Two class SVM classification using kPCA

SVM using kPCA	Least Squares	Sequential minimal optimization	Quadratic programming
Linear	81.4433	81.4433	81.4433
Polynomial	78.3505	79.3814	79.3814
RBF	82.4742	85.5670	82.4742
Quadratic	79.3814	80.4124	-
Multilayer perceptron	51.5464	74.2268	-

In two class classification the feature reduction with most widely used PCA and also the kernel PCA were compared. The table 1 depicts the SVM classification over various kernels without using PCA or kPCA. Table 2 depicts the SVM classification over various kernels using PCA and table 3 depicts the SVM classification over various kernels using kPCA. The above observation clearly shows that PCA works well for linear and multilayer perceptron kernels while kPCA gives a far better accuracy for polynomial, radial basis function and quadratic kernel than our widely used PCA.

Table IV Multiclass SVM classification without feature reduction

SVM without PCA or kPCA	Least Squares	Sequential minimal optimization	Quadratic programming
Linear	71.0246	74.9205	70.7669
Polynomial	53.3253	59.6436	59.6436
RBF	61.4636	61.7213	61.4636
Quadratic	55.6371	58.3774	-
Multilayer perceptron	52.9578	67.0417	-

Table V Multiclass SVM classification using kPCA for feature reduction

SVM using kPCA	Least Squares	Sequential minimal optimization	Quadratic programming
Linear	90.9515	94.5876	94.5876
Polynomial	81.7690	81.1533	80.6379
RBF	63.6256	64.5590	63.6256
Quadratic	76.4703	75.4122	-
Multilayer perceptron	56.9359	83.1599	-

We also have tried our kernel PCA technique on the multi class classification where we have five classes namely 0, 1, 2, 3, 4 by using *one-versus-the-rest* approach given by Vapnik[18]. In multiclass classification SVM classified with much better accuracy when the features were reduced using kPCA. The emphasis on why we considered multiclass classification was, there can be many types of heart diseases emerging from the same data and they need to be classified promptly for a better accuracy. We have done the experiment without using the feature reduction and also using kPCA for reducing the features and it gave a better accuracy than that of our conventional PCA.

Conclusions

From this study we can infer that though PCA is widely used feature reduction technique, kPCA has far better accuracy especially when the data is non-linear because it could make the non-linear data into a linear one by taking it to a higher dimensional plane and then applying the normal PCA.

References

[1] Avci, E.; Turkoglu, I., "An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases". Journal of Expert Systems with Application, Vol. 2, No. 1, 2009, pp. 2873-2878.

[2] Bailin Liu, Huiyun Zou, Xi Chen, "*Disease Diagnose based on Support Vector Machine and Rough Set*," in press.

[3] Baliga, R. R., & Eagle, K. A. (2008). "Practical cardiology: Evaluation and treatment of common cardiovascular". Lippincott Williams & Wilkins.

[4] Christopher M. Bishop, Pattern Recognition and Machine Learning., pp. 586-590, Springer, ISBN-10: 0-387-31073-8.

[5] Comak, E., Arslan, A. and Ibrahim, T. (2007).'A decision support system based on support vector machines for diagnosis of the heart valve diseases'. *Computers in biology and Medicine* (37), pp 21-27.

[6] Diagnosis, E. (2010). Coronary disease or heart attack from expert system: Chest pain [Online]. Available: <http://www.easydiagnosis.com/cgi-bin/expert/ [Accessed 12th December 2010].

[7] Fan, Y., & Chaovalitwongse, W. (2010). 'Optimizing feature selection to improve medical diagnosis'. Annals of Operations Research, 174, 169–183.

[8] Huang, C.-L. and Wang, C.-J.,'A GA-based feature selection and parameters optimization for support vector machines', *Expert Systems with applications*, Vol.-31,(2006), pp -231-240.

[9] Jesmin Nahar a, Tasadduq Imama, Kevin S. Tickle, Yi-Ping Phoebe Chen, "Computational intelligence for heart disease diagnosis: A medical knowledge driven approach", Journal of Expert Systems with Applications, 40, 2013, pp. 96–104.

[10] K.R. Lakshmi, M.Veera Krishna and S.Prem Kumar (2013) 'Performance Comparison of Data Mining Techniques for predicting of Heart Disease Survivability', International Journal of Scientific and Research Publications, Volume 3, Issue 6, June 2013.

[11] Mi Hye Song, Jeon Lee, Sung Pil Cho, Kyoung Joung Lee, Sun Kook Yoo, "Support Vector Machine Based Arrhythmia Classification Using Reduced Features", International Journal of Control, Automation, and Systems, vol. 3, no. 4, pp. 571-579, December 2005.

[12] Pickett, J. P. (2000). The American heritage dictionary of the english language. Houghton Mifflin.

[13] Scholkopf, B., A. Smola, and K.R. Muller (1998).Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation* **10**(5), 1299–1319.

[14] Soman, K. P., Shyam, D. M., & Madhavdas, P. (2003). 'Efficient classification and analysis of ischemic heart disease using proximal support vector machines based decision trees, TENCON'. In Conference on convergent technologies for Asia- Pacific region (Vol. 1, pp. 214–217).

[15] Sumit Bhatia, Praveen Prakash, and G.N. Pillai, "SVM based Decision Support System for heart disease classification with Integer-Coded Genetic Algorithm to select critical features" Proceedings of the World Congress on Engineering and Computer Science 2008 WCECS 2008, October 22 - 24, 2008, San Francisco, USA.

[16] Uci. 2010. Cleveland heart disease data details [Online]. Available: <http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/heartdisease.names> [Accessed 8th February 2010].

[17] V. Vapnik. The Nature of Statistical Learning Theory. NY: Springer Verlag. 1995.

[18] Vapnik, V. N. (1998). *Statistical learning theory*.Wiley.

[19] Yan, H., Zheng, J., Jiang, Y., Peng, C. and Xiao, C., ' Selecting critical clinical features for heart diseases diagnosis with a real-coded genetic algorithm',Applied Soft Computing, Vol.-8,(2008), pp 1105-1111.

[20] Yang J. and Honavar, V., 'Feature Subset Selection using a Genetic Algorithm', *IEEE Intelligent Systems*, March-April 1998, pp 44-49.

[21] Zhao, H., Chen, J., Hou, N., Zheng, C., & Wang, W. (2010). 'Identifying metabolite biomarkers in unstable angina in-patients by feature selection based datamining methods'. In Second international conference on computer modeling and simulation (pp. 438–442). Sanya, China: IEEE.