International Journal of the Physical Sciences Vol. 6(33), pp. 7540 - 7549, 9 December, 2011
Available online at http://www.academicjournals.org/IJPS
DOI: 10.5897/IJPS11.469
ISSN 1992 - 1950 ©2011 Academic Journals

Full Length Research Paper

A subspace ensemble based data dependent binary classification model

Muhammad A. Khan[1]* and Anwar M. Mirza[2]

[1]Department of Computer Science, National University of Computer and Emerging Sciences
A. K Brohi road, H-11/4, Islamabad 44000, Pakistan.
[2]College of Computer and Information Sciences, King Saud University, Saudi Arabia.

Accepted 24 June, 2011

Classifying patterns into two classes is a typical problem of binary classification in pattern recognition. Binary classification is an industrial problem in many fields like medicine, search mechanism, diagnostic of disease in humans, security and many other aspects. In this paper, we have proposed a random subspace based ensemble data dependent classification model for the binary classification problem. The proposed method makes use of the information about the structure of given data and the availability of the training instances, in selection of the classification model. A subspace ensemble for a set of one class and two-class classifier are generated, trained and tested on the given data. The proposed method is evaluated on receiver operating curve (ROC), cross validation accuracy and Q-statistics. From the empirical results obtained, we have reached the following conclusions that the overall performance of the two class ensemble was better because of the ability of the ensemble to make use of the knowledge of both the positive and negative samples and thus constructs better class boundaries. The one class ensemble makes use of positive samples only and gives better performance when (i) training patterns are sparse and (ii) outlier detection is required.

Key words: One class classifier, two-class classifier, binary classification, classification model, receiver operating curve (ROC), evaluation, Q- statistics.

INTRODUCTION

In the literature, the pattern recognition community has proposed cost effective, robust and efficient solutions for binary classification problem. Majority of the proposed classification system performs in some constrained environment; the need to discover new ways of pattern classification is always needed and welcome in the community. In binary classification problems, we have objects belonging to two categories or groups and a corresponding category or group for a new previously unseen pattern has to be determined. The objects in a dataset Z are represented by an n-dimensional feature vector $x=[x_1, x_2, ..., x_n]^T \in \Re^n$ and a set of two given classes $\Omega=\{\omega_1, \omega_2\}$. In binary classification, each object classification problems, the application of a model must not be treated as a matter of random choice instead the selection of classification model must be based on the detailed study of the model and the pertinence of the model to be applied in that particular problem (Bishop, 1996).

Search for ideal pattern recognition is on in the research community, yet there is no such system which can be claimed so. Practically, in solutions for classification problems, we have to limit ourselves to a tradeoff. For example, achieving 100% accuracy and avoiding the chance of over fitting and over training of the classification model, is a scenario where we compromise on one of the parameters. The research questions that are addressed in this paper are: (1) In which cases either of the classification models (multi-class or one-class) is better performing? (2) What is the plausible explanation of the better performance of the multi-class classifier on

*Corresponding author. E-mail: aamir.khan@nu.edu.pk.

the dataset? (3) What is a suitable measure to evaluate the models to have a fair and objective comparison (Tax and Duin, 2001)?

The major contribution of the paper is described as follows: (i) The proposed method creates an ensemble based classification model depending on the data given in a particular problem, a data dependent classification model; (ii) The parameters and the constraints for ensemble creation and its better performance are presented; (iii) The evaluation of the proposed method is based on the diverse and sound evaluation measures such as receiver operating curve (ROC), cross validation and statistical measures such as Q statistics.

LITERATURE REVIEW

One of the most researched and experimented field of study in machine learning society is pattern classification. In pattern classification for each pattern in the dataset *Z*, a classifier assigns an output class label to it. From the learning perspective and availability of class labels classification algorithms are divided into two major categories namely (1) unsupervised learning and (2) supervised learning.

This paper tries to solve the problems and short comings in pattern recognition associated with supervised learning domain. In supervised learning, there is a selection of classifier, training followed by testing of the classification model. However the study can be extended to unsupervised classification too. Subsequently, a brief description of one-class and two-class classification model is presented.

One-class ensemble based classification

In one class classification problem, the description of a particular class called as target class is learned by the algorithm. New patterns are classified according to that learnt description of the target class. The objects are assigned the label of the target class if correctly classified otherwise the patterns that do not belong to the target class as labeled as outliers (Tax, 2001), that is the reason one-class classification is also referred as outlier detection. One class classification model focuses on the problems of classification where a well defined training pattern is available for one of the target classes. In many cases, a one-class classifier is used in preference to a multi-class classifier because of either of the reasons, (1) it may be difficult to use non-target data in training or (2) only data of a single category is available. Some applications of one-class classification are machine monitoring, password hardening (Cho, 2003), typist recognition (Nisenson, 2003), novelty detection, image database retrieval and authorship verification (Schler, 2004). One class classification has application in medical problems as well such as tumor detection, where a limited quantity of negative data is available during training process (Tarassenko et al., 1995).

Mathematically, a one class classifier is expressed as

$$h(x) = \begin{cases} target \;\; if & f(x) \le \theta \\ outlier \;\; if & f(x) > \theta \end{cases} \quad (1)$$

where θ is a threshold calculated according to the maximum allowable error on target class defined.

For example, the gender classification problem is a typical binary classification problem, assume for training the female subjects data is available, a one class classifier *x* will be trained on the female subjects data and the female class will be the target class of this classifier. The new incoming patterns in the testing phase will be classified on the basis of the description learned from target class. The new patterns will be assigned the label of the female class (target class) or they will be classified as outlier (any pattern not belonging to the target class will be assigned the label outlier).

Two-class classification model

The classifier learns from the training pattern of both the classes in the two class classification problem. The boundary of the two classes is constructed from the description extracted from the training patterns of the two classes. The two class classification is essentially a multi classification model with two classes under study. Decision trees, support vector machine (SVM) and neural network (NN) are some of the famous and most studied two class classifiers. Support vector machine (SVM) is developed by Vapnik at bell labs. SVM is a margin classifier; the decision boundary is defined by the function $f(x) = \mathbf{w}\,\phi(\mathbf{x}) + b$ depending on the sign of the function the pattern x is classified into either of the two classes. Neural networks are modeled after biological neurons. Perceptron is a feed forward neural network implemented as, if $u = [u_0, \ldots, u_q]^T \in \Re^{q+1}$ be input vector to the node, $v \in \Re$ be its output and $w = [w_0, \ldots, w_q]^T \in \Re^{q+1}$ be a vector of synaptic weights and an activation function ϕ then $v = \phi(\xi)$ where ξ the sum of weights is $\xi = \sum_{i=0}^{q} w_i u_i$. There are many types of activation functions namely sigmoid, threshold and identity functions. In perceptron, a threshold activation function is used $\phi(\xi) = \begin{Bmatrix} 1, & if\ \xi > 0, \\ -1, & otherwise \end{Bmatrix}$.

For problems with more than two classes, to classify high level multi-class methods have been developed which uses the two class classifiers as the basic building blocks. The multiclass methods are (1) one-against-all and (2) one-against-one; these methods are briefly described as follows.

One-against-All

In one-against-all strategy, one classifier is constructed per class and trained to distinguish the samples of one class from the samples of the remaining classes. A new pattern is assigned a class label of the classifier which gives the maximum accuracy among all other classifiers.

One-against-one

This strategy for multi class classification is also called *"pair wise coupling".* For C classes problem, $\frac{c(c-1)}{2}$ classifiers are trained to distinguish sample of one class from sample of another class. A pattern is assigned class label of the classifier based on the majority voting aggregation rule.

Ensemble approach

Machine learning community has proposed to use combination of multiple classifiers to improve the performance of pattern recognition systems in comparison to the single classifier models. The multi classifier systems are based on the combination of a pool of classifier with the aim that the fusion of their decisions yields higher classification accuracy as compared to single classifier. Ensemble of classifier is a technique where a pool of classifiers is trained and their decisions are combined through certain fusion rule. Some of the famous and most practical ensemble techniques in the pattern recognition literature are bagging (Breiman, 1996), boosting (Breimen, 1993), random subspace (Ho, 1998), class switching (Suarez, 2005) and rotation forest (Rodriguez et al., 2006). In this paper, we have used the random supspace method (Ho, 1998); in this ensemble, each classifier created is independent of the other and it works like a prallele learning algorithm. In the literature, it is emphizised that the random supspace ensembel avoid hill climbing phenomena and is reslilient to sticking in a local minima. Generally, the random subspace is applied to the decision trees but studies show that it had been used to generate ensembles of classifiers other than decision trees. However, in the random subspace of the knn classifier, it was shown to have better performance (Alkoot and Kittler, 2001).

The basic principle for constructing a random subspace ensemble is illustrated in the following steps

1. A random projection of the d dimensional feature space to a k- dimensional subspace is selected
2. Project the data into the k- dimensional subspace
3. Train the selected either one class or two class classifier on the selected sub space projection
4. Repeat steps 1 to 3 m times to generate m- different subspaces
5. Combine the classifier decision using the Majority Voting aggregation rule.

Related work

Hempstalk et al. (2008) in their work used one class classification model in cases where negative data is not available or the collection of negative data is impossible. One class model also helps in scenarios when there are new classes that were not available at the training times means classes that are not previously seen by the classifier (Tax, 2001). Their work has establish the robustness of one class classification models to outliers, with the prior knowledge of known outliers if incorporated further help tighten the boundaries of the target class. The training cost of one class classifier is less as compare to the two class classification model. Three approaches for one class model creation have been described in this work. Kennedy (2009) have used the one class classification model for the credit scoring and presented a detailed review of comparison of the one class classification models with two class, the generation of classification model and evaluation of the classification mode. In their work, they have also presented the work done by other researchers in different credit scoring domains using the one class classification. Barandela (2003) describes the class imbalance problem as the number of sample belonging to one of the classes in the training sample is represented by a very small number as compare to the other classes; class imbalance is found in such training sets.

In their work, Weiss (2004) presented problems that arise with class imbalance in datasets and also the possible solution to the class imbalance problem. Sampling, boosting and cost sensitive learning are one of the solutions proposed for the class imbalance problem.

Random subspace has widely been used for classification problems. It is an established fact in the pattern recognition community that the ensemble of classifier improves the discriminative performance of the classification system. Random subspace was proposed by Ho (1998). It had been modified for performance and used in an array of different areas of classification. One improvement suggest by x is the introduction of re-sampling in the ensemble generation phase (Yanping, 2010). Re-sampling helps create weak classifiers which in turn will result in a stronger aggregation result. Here, the related work in random subspaces from the view of class imbalance, binary classification and one class classification is presented.

PROPOSED METHOD

The method used to select a particular classifier model for a said

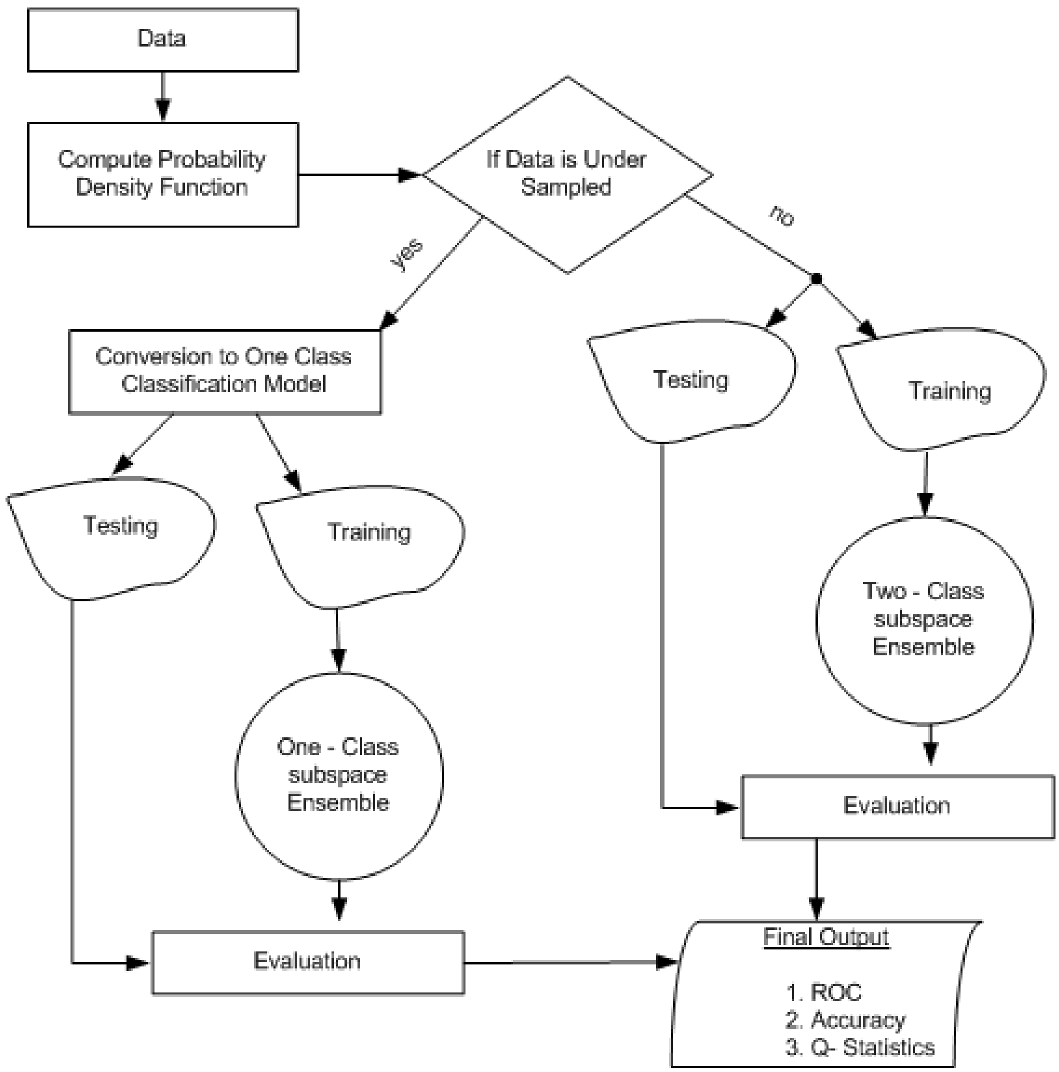

Figure 1. Flow chart of the proposed method for selection of classification model.

dataset is described here. In the proposed method, the distribution of classes in the data is known with the help of probability density function. Based on the class distribution, the classification model for training is decided. If the distribution exhibits that there are enough patterns from both the classes which dictate the notation that there is sufficient amount of negative as well as positive samples in the data then, two class classification model is selected for training. The two class classification model exploits the data from both classes. Conversely, if the class distribution reveals that imbalance is present in the class distribution then one class classification model is selected for training. The class with the maximum frequency is selected as the positive data for training with the one class classifier; this is a step of conversion of the dataset into one class form. The dataset is divided into training and testing set, while the ration is kept as 70% training and 30% testing. In one of our earlier studies (Muhammad et al., 2009) on classifier fusion and rotation of the datasets, we found this ratio of train to test as one of the best possible ratio. Classifiers from either the models are trained on the train set and then evaluated on the test set. We decided to use receiver operating curve, Q statistics and Cross validation accuracy as the performance measures; they are described in detail in the result and discussion.

In Figure 1, the flow chart of the method is given. A set of objects or pattern is presented as input data. The proposed algorithm tries to find the underlying structure of the classes in the data with the help of probability density function. Prior knowledge of probability distribution comes handy in the ‘classification model’. This information is a kind of prior knowledge in the scenarios when the input data is increasing with time or new data is generated like in case of data streams then finding class distribution cannot be computed before the classification stage. This is the main limitation of our proposed method. Table 1 shows the class distribution of the six bench mark datasets.

The result of the random subspace ensemble is aggregated using the majority aggregation rule. The result of the one class and the two class classifier are aggregated and a final decision is obtained.

EXPERIMENTS

Experimental setup

We used four dataset from the UCI data set repository (Asuncion

Table 1. Class distribution of patterns in two classes.

Data set	Percentage of patterns in classes	
	Class 1	Class 2
SUMS	50	50
WDBC	62.74	37.25
WPBC	76.26	23.73
Ionosphere	45.58	54.41
German credit	37	63

Table 2. List of the datasets used in the experiment.

Dataset	Class	Instance	Dimension
SUMS	2	400	25
WDBC	2	569	30
WPBC	2	198	32
Ionosphere	2	351	34
German credit data	2	1000	34

Table 3. List of multi-class classifier used in the experiments.

Classifier name	Description
K- Nearest neighbors	With 3 nearest neighbors
QDC	Quadratic bayes normal classifier
Linear perceptron	A neural network
Support vector machine	With RBF kernel

and Newman, 2007) and one from the Stanford University Medical Student Dataset. All the experiments are coded in MATLAB 7.3 on Intel core 2 duo machine with 1 GB of RAM.

Data sets

The dataset were used in a ratio of 65% for training and the remaining 35% for testing. In the WPBC, four records contain missing values for certain attributes. All the four records were eliminated from the dataset. The Data sets used in the experiments are listed in Table 2.

Multi-class classifier

We have used four different multi class classifiers listed in Table 3. Classifiers are of diverse nature and suited for binary classification. All the classifiers are briefly described following.

KNN is a well known classifier which takes into account the distance of the neighboring objects. K-NN classifier falls under the supervised learning category of classifiers. K-NN classifies the unseen patterns by assigning it the class label most frequently represented by its k nearest neighbors. The performance of the K-NN classifier depends on two measures namely, (1) the distance between the patterns and (2) the optimal value of k neighbors (Kotsiantin, 2007).

In scenarios where classes are normally distributed, Quadratic classifier (QDC) is the optimal classifier for the problem. In most cases, we do not have sufficient knowledge of the underlying distribution still the QDC classification model yields reasonable accuracy. Quadratic classifier (QDC) derives its name from the type of discriminant function it uses. The data for the QDC classifier is assumed to be normally distributed with specific class covariance matrices. For the class labels under study, a set of discriminant functions are obtained.

$$gi(\mathbf{x}) = w_{i0} + \mathbf{w}_i^T \mathbf{x} + \mathbf{x}^T W_{i\mathbf{x}} \qquad (2)$$

where $gi(\mathbf{x})$ a set of discriminant function, and $w_{i0} \in \Re$, $\mathbf{w}_i \in \Re^n$ are the coefficients of the discriminant function $gi(\mathbf{x})$.

Linear perceptron is a neural network with proved classification performance. Support vector machine (SVM) is a maximum margin classifier and regarded as one of the best binary classifier in literature. SVM with RBF kernel is used in our experiments.

One class classifier

The one-class classifiers used in the experiment are listed in Table No. 4. The Gaussian data descriptor classifier models the target class as a Gaussian distribution based on the Mahalanbolis distance

Table 4. List of one class classifier.

Classifier name	Description
Gaussian data description	Based on Gaussian distribution
K means	Standard K means clustering
SVDD	Support vector data description
Knndd	One class knn classifier

(Duda et al., 2001). Equation 2 represents a Gaussian data descriptor

$$f(x)=(x-\mu)^T \sum^{-1}(x-\mu) \quad (3)$$

where μ is mean and $\sum$ is the covariance matrix.

K-means classifies the data into K clusters by means of standard K-means clustering procedure (Bishop, 1996). The K-mean is a classifier based on the statistical measure mean. Division of data in k-mean is solely based on the natural separation of the data using mean values. The algorithm works by picking K, the number of groups that we intend to create. The Euclidean distance from each training pattern to each data center is computed. Each training pattern is associated with the data center point which is closest to it. After the binding of each training pattern with their respective closest data centre, a new data center is calculated for each group based on the mean value of all the data points in that group. This process is repeated iteratively until mean value changes no more. Grouping data based on the Euclidean distance between a test point and two center points is same as dividing the data with a hyper-plane that splits the two center points. The average distance of data objects x to the cluster center c is minimized. The target class is defined as

$$f(x)=\min(\mathbf{x}-\mathbf{c_i})^2 \quad (4)$$

Support vector data descriptor (SVDD) (Tax, Support Vector Domain Descriptor, 1999, 2004) tries to fit a hyper sphere around the target class boundary. All the objects inside the hyper sphere are target objects and the objects outside the hyper sphere are deemed as outliers. The hyper sphere can be optimized with application of kernels; some of the famous are kernel whitening, radial basis function, etc.

Knnd is based on the density estimation computed by a simple nearest neighbor classifier. It uses the distance estimation among pattern in the feature space and avoids explicit density estimation. A pattern x is classified as true when its local density is larger or equal to the density of its nearest neighbor in the training set.

RESULTS AND DISCUSSION

The fundamental difference between one class and two class classifier is the utilization of the negative data instances in training by the two class classifier function. The one class classifier approach has the advantage of using a smaller training set, less space and lesser training time. In some problems, there exist a large amount of known data and it is not desirable to use all the data in training or we may not even know the relevant data in such problems; only the data of the class to discover is used. The following are the evaluation results of the classifiers presented with respect to the measure used.

Receiver operating curve (ROC)

The receiver operating curve (ROC) performance of one class and two class classifier is presented on all of the five bench mark data sets. ROC is a preferred technique of comparing classification performance of two classifiers. ROC is a two dimensional measure of classification performance that plots the true positive rate (TPR) against the false positive rate (FPR). The true positive rate (TPR) is defined as the ratio of the number of correctly classified positive cases to the total number of positive cases. Mathematically, TPR is expressed as,

$$TPR=\frac{TP}{TP+FN}$$

where TP stands for *true positive* and FN stands for *false negative*

The false positive rate (FPR) is defined as the ratio of incorrectly classified negative cases to the total number of negative cases. Mathematically, FPR is expressed as,

$$FPR=\frac{FP}{FP+TN}$$

where FP stands for *false positive* and TN stands for *true negative*.

The diagonal line y = x corresponds to a classifier which predicts a class by random guessing; A classifier having ROC curve above this line is regarded as useful classifier. An optimal classifier is the one that generates a curve in the upper left corner of the ROC space This means that the classifiers has predicted all the classes accurately making no mistake. (Fawcett, 2004) Detailed and thorough observation reveals that the multi class classifier exhibits better performance ROC on all the datasets as compare to the one class classifier.

Two-class classifier takes into account the ‘negative data’ or the data about the *‘non target’* class which make it able to classify the objects with a broader knowledge about both classes (Tax and Duin, 2001). The ROC of

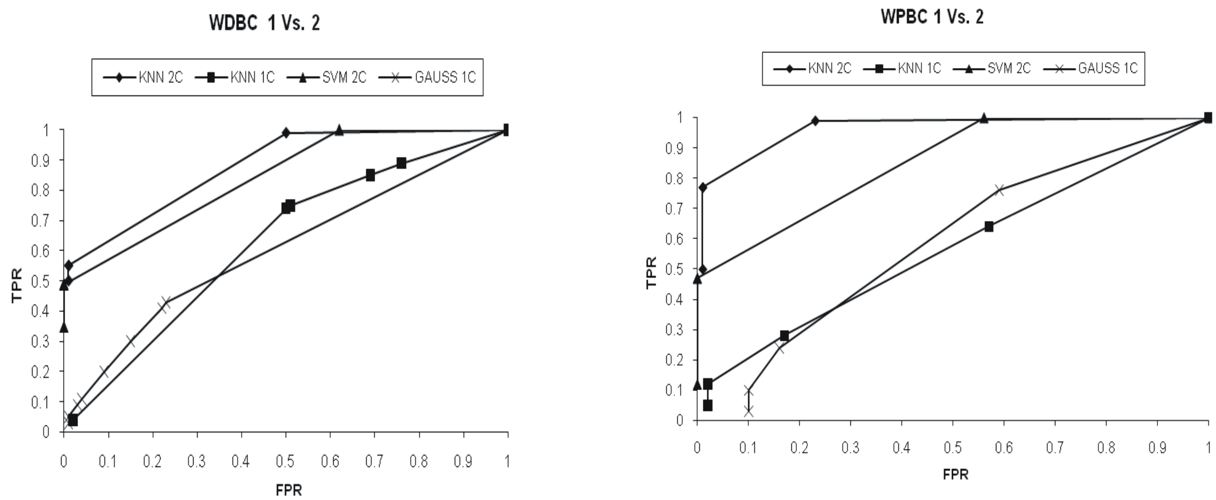

Figure 2. Receiver operating curve of WPBC and WDPC on one class and two class classifiers.

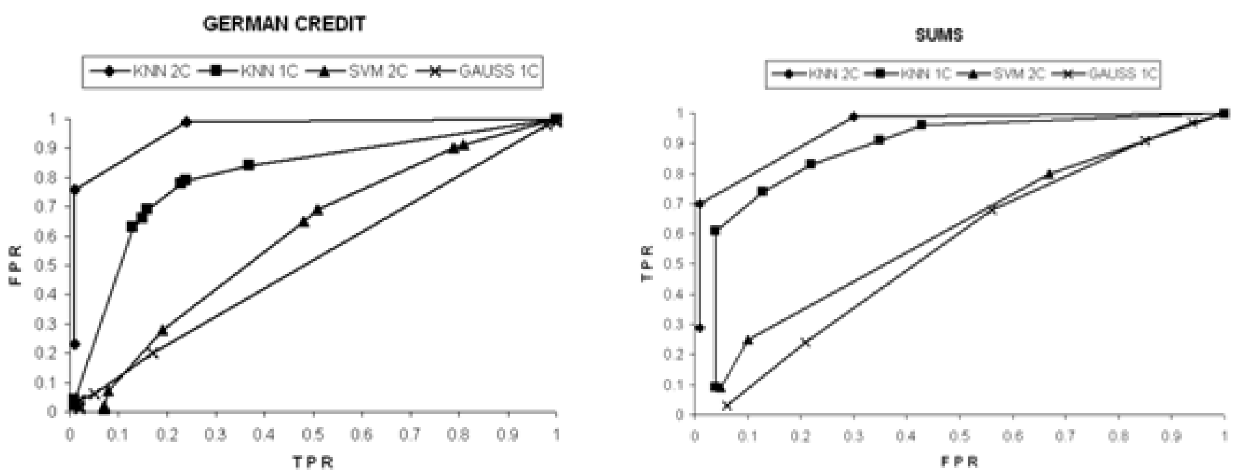

Figure 3. Receiver operating curve of one class and two class classifier on the German Credit and the Stanford Medical Student Dataset (SUMS) for gender classification.

some of the classifier on the six dataset is presented in Figures 2, 3 and 4.

Q Statistics (Pair wise measure)

Q statistics for one class classifier is in Table 6 and the value of Q-statics for multiclass classifiers is given in Table 5. For two classifiers D_i and $D_{k,}$ the Yule's Q-statistics (Kuncheva, 2003) is expressed as

$$Q_{i,k} = \frac{ad - bc}{ad + bc} \tag{5}$$

where *a* is the probability of both classification models being correct, *d* is the probability of both the classifier being incorrect, b is the probability that first classifier is correct and second is incorrect, c is the probability second classifier is correct and first is incorrect. Value of Q varies between -1 and 1. Classifier that tends to classify the same patterns correctly will have a value of Q > 0 and the classifiers that gives erroneous results on different patterns will have value of Q <0. The average value of Q is calculated as shown in (Kuncheva, 2003). For a pair of classifiers, the average Q values is calculated as follows

$$Q_{av} = \frac{2}{L(L-1)} \sum_{i=1}^{L-1} \sum_{k=i+1}^{L} Q\,i,k \tag{6}$$

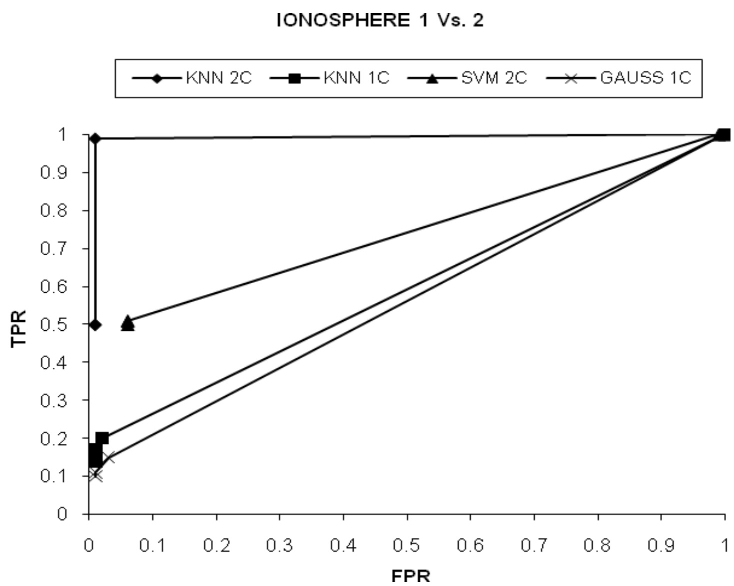

Figure 4. Receiver operating curve of ionosphere for the one class and two class classifiers.

Table 5. List of Q-statistics value for the two-class classifiers.

Classifier model	WPBC	WDBC	German Credit	SUMS	Ionosphere
KNN	0.5042	0.9824	0.5810	0.8271	0.9863
QDC	0.6979	0.9909	0.5589	0.8692	0.9933
Perceptron	0.7636	0.9903	0.6406	0.8066	0.9944
SVM	1	1	0.6157	0.9180	0.9792

Table 6. List of Q-Statistics value for the one-class classifier.

Classifier model	WPBC	WDBC	German Credit	SUMS	Ionosphere
Gauss_dd	0.8255	0.9857	0.8887	0.9670	1
Kmeans	0.8689	0.9874	0.9175	0.9619	0.9986
Svdd	1	1	1	0.9786	0.9986
Knnd	0.9565	0.9983	0.9175	0.9694	1

Cross validation error/accuracy

Cross validation errors are more significant than accuracy alone. In the training phase, the classification model tends to over fit on the training data. Over fitting is major problem in limiting the generalization capability of any classification model. To avoid the pitfall of over fitting, we have reported the tenfold cross validation error. Cross validation error are reported for both the classification models in Tables 7 and 8.

Conclusion

In this paper, we presented a method to select a classification model for a given data set based on the lessons learned from the application of two-class and one-class classification model to binary classification problem. One-class classification model is good at outlier detection and in scenarios when only the training data of the target class is available. Two-class classification model are more versatile and they construct the class boundaries using the information from the training data of both the classes. Our proposed method incorporates the benefits of the two classification model and automatically selects the suitable classification model for the dataset depending on the underlying structure of the dataset. We have used a diverse pool of base classifiers for the classification purpose and adapted the single best

Table 7. Cross validation error for the two-class classifiers.

Classifier model	WPBC	WDBC	German Credit	SUMS	Ionosphere
KNN	0.2319	0.0738	0.303	0.4475	0.135
QDC	0.2319	0.0456	0.231	0.4350	0.132
Perceptron	0.2938	0.0281	0.293	0.4675	0.144
SVM	0.2268	0.0492	0.231	0.4750	0.125

Table 8. Cross validation error for the one-class classifiers.

Classifier model	WPBC	WDBC	German Credit	SUMS	Ionosphere
Gauss_dd	0.1090	0.5014	0.2400	0.3750	0.7607
Kmeans	0.6539	0.4114	0.2500	0.4750	0.7607
Svdd	0.1186	0.2839	0.4200	0.5000	0.7009
Knnd	0.7084	0.4836	0.2100	0.4250	0.6709

classifier approach for the classification output. Overall, the performance of the two class classification model was better than one class model on the six datasets. The plausible explanation is that due to the knowledge of both the classes, the two-class model achieves better performance but in case of unavailability of sufficient data; the one class model is preferred. In future, we aim to apply ensemble approach to the classification models with optimization of the base classifier through evolutionary and hybrid methods.

ACKNOWLEDGEMENTS

The authors would like to thank the Higher Education Commission (HEC) for their funding the Indigenous PhD Scholarship Program. This work was supported by the Research Center of College of Computer and Information Sciences, King Saud University. The authors are grateful for this support. The authors also extend gratitude for the financial support of Institute of Management Sciences (IM|Sciences) Hayatabad, Peshawar.

REFERENCES

Alkoot FM, Kittler J (2001). Population bias control for bagging k-NN experts. Processdings sensor fusion: Achtectures, algorithm and Appication. 23(8).

Asuncion A, Newman DJ (2007). UCI Machine Learning Repository. University of California, Shcool of Information and Computer Science. 4637: 178-197.

Bishop C (1996). Neural Networks for Pattern Recognition. Oxford University Press, Walton Street, Oxford OX2 6DP.

Breiman L, Friedman JH, Olshen RA, Stone CJ (1993). Classification and Regression Trees. Chapman & Hall, New York, 358 p.

Breiman L (1996). Bagging Predictors. Machine Learning, pp. 123-140.

Cho EY (2003). Novelty detection approach for keystroke dynamics identity verfication. International Conference on Intelligent Data Engineering and Automated Learning. Berlin: Springer, pp. 1016-1023.

Duda RO, Hart PE, Stork DG (2001). Pattern Classification, 2nd Edition. New York: John Wiley & Sons.

Fawcett T (2004). ROC Graphs:Notes and practial consideration for Researchers. Palo Alto, USA: HP Laboraties. Available at http://www.purl.org/NET/tfawcett/papers/ROC101.pdf.

Hempstalk K, Eibe F, Ian HW (2008). One-class classification by combining density and class probability estimation. Machine Learning and Knowledge Discovery in Databases. Springer, 5211: 505-519.

Ho TK (1998). The random subspace method for constructing decision forests. IEEE Transaction of Pattern Analysis and Machine Learning, 20(8): 832-844.

Kennedy KA (2009). Low-Default Portfolio/One-Class Classification: A Literature Review.

Kotsiantin S (2007). Supervised Machine Learning: A review of Classification Tehnique. Informatica , 31: 249-268.

Kuncheva LW (2003). Measures of Diversity in classifer ensemble and their relationship with ensemble accuracy. Machine Learning, 51: 181-207.

Lee HM, Chen CM, Chen JM, Jou YL (2001). An Efficent Fuzzy classifier with Feature selection based on Fuzzy Entropy. IEEE Transactions on System, man and Cybernetics-PARTB: CYBERNATICS, 31(3): 426-432.

Martinez-Munoz G, Suarez A (2005). Switching Class labels to generate classification ensembles. Pattern Recognition, 38(10): 1483-1494.

Muhammad AK, Zahoor J, Anwar MM (2009). Performance Analysis of Classifier Fusion Model with minimum Feature Subset and totaion of the dataset. Fuzzy System and Knowledge Discovery (FSKD) Tianjin: IEEE, pp. 251-255.

Nisenson M, Yariv I, El-Yaniv R, Meir R (2003). Towards behaviometric security systems: Learning to identfy a typist. 7th International conference on Principles and practices of Knowledge Discovery in Databases Berlin Germany: Springer-Verlag, pp. 363-374.

Rodriguez JJ, Kuncheva LI, Alonso CJ (2006). Rotation Forest: A new classifier ensemble method. IEEE Transaction of Pattern Analysis and Machine Learning , 28(10): 1619-1630.

Schler J, Koppel M (2004). Authorship verification as a one-class classification problem. 21st International Conference on Machine Learning Newyork, NY, USA: ACM Press, pp. 489-495.

Tarassenko L, Hayton P, Brady M (1995). Novelty detection for identification of masses in mammograms. Fourth International IEEE Conference on Artificial Neural Networks London, England: IEEE, pp. 442-447.

Tax DMJ, Duin RPW (2001). Combining One-Class Classifiers. Multiple Classifier Systems MCS. Cambridge, UK: Springer-Verlag, 2096: 299-308.

Tax DD (1999). Support Vector Domain Descriptor. Pattern Recognition Letters, 20: 11-12, 1191-1199.

Tax DM (2001). One Class Classification, Concept Learning in the Absence of Counter Example. Delft, Netherland: Delft University of Technology, pp.1-190.
Tax DD (2004). Support Vector Domain Descriptor. Machine Learning, 54(1): 45-66.
Weiss G (2004). Mining with rarity: a unifying framework. ACM SIGKDD Explorations Newsletter, 6(1): 7-19.
Yanping Yang GM (2010). Ensemble - based active learning for class imbalance problem. J. Biomed. Sci. Eng., 3: 1021-1028.