

Kernel Collaboration with Perceptron for Facial Emotion Recognition

R.Bhuvaneswari
Assistant Professor
Sri Venkateswara College of Engineering
Sriperumbudur

K.Thaiyalnayaki
Associate Professor
Sri Venkateswara College of Engineering
Sriperumbudur

ABSTRACT

Many classifiers are in existence for handling linear and nonlinear type of data. In this paper, a method for handling nonlinear data for classification is proposed. The classifier is modeled for both two-class and multi-class problem. In this framework, the input samples are projected into a higher dimensional feature space by utilizing kernel trick. In order to avoid the difficulty of working in this higher dimension space, a dimensional reduction technique is used. This reduction technique results with the transformation matrix which is of lesser dimensions. Then the classification of data is carried out in this reduced subspace by using perceptron technique. The classifier is utilized for recognizing the emotion of a person based on the facial expressions. Experiments are carried on toy data set and JAFFE data set and also the results have shown the effectiveness of this classifier. The classifier recognizes the 6 different Emotions with 98.6% efficiency.

General Terms

Pattern Recognition, Classification, Linear and Non linear data.

Keywords

Kernel trick, perceptron, dimension reduction technique, classification.

1. INTRODUCTION

Pattern Analysis is the measure of finding general relations in a dataset and it forms the basis for many applications which ranges from medical diagnosis to defense. Pattern analysis plays an important role in gathering the knowledge from the raw data. As the modern computing technologies is handling large amount of data, different techniques are needed for extracting the hidden information from it. Given a set of data which belongs to two classes, the classifier has to recognize the class for each dataset. The final and ultimate goal of the classifier is to classify the given input patterns to known classes. When the class information is available the classifier is trained with the training samples which is termed as supervised learning. When the class information is not available then the datasets are grouped into a class under the similarities between them. This method is termed as unsupervised learning.

Though, the sample data belongs to different class distribute on the same vector direction, the classifier classifies the multiclass datasets properly. The data has been mapped on to a higher dimensional space in order to make it linearly separable in this input space. A more recent method of constructing kernel algorithm where the starting point is not a linear algorithm but a linear criterion. Nonlinear data are handled using any of the kernel technique.

This project aims at classifying the nonlinear dataset with high accuracy. The nonlinear data are projected in kernel space using any one of the kernel tricks. Then the data are classified using perceptron technique where the weight vectors are adjusted while training the classifier until it is classified correctly.

The remainder of this paper is structured as follows:

Section II discusses related work. Section III presents the working of classifier, kernel collaborative with perceptron. Section IV presents the accuracy results of classifier and application of this classifier for emotional dataset. Finally, concluding remarks are given in section V.

2. RELATED WORK

Designing the perfect classifier is often impossible. The difficulty in the classification lies on the variability in the feature values for objects in the same category relative to the difference between objects. The task of the good classifier is to assign the object to a class by using the feature vector. The simple measure of classifier performance is the classification error rate – the number of new patterns that are assigned to the wrong class (i.e misclassified).

Among linear and nonlinear classifier, perceptron is the simplest form a neural network used for classification. Perceptron consists of a single neuron with adjustable weights and bias. The operation of perceptron is based on error-correlation learning. The learning process for pattern classification occupies a finite number of iterations and then it stops.

According to [1] the classifier shows the good performance results on face image data in which they combined the result of machine learning and compressed sensing. This paper overcomes the drawback of sparse representation classifier. The input feature is mapped into a high-dimensional kernel feature using the kernel trick. As there exists complexity in working in this higher dimension, a dimension reduction technique is used. They indeed reduce the dimensionality of the kernel feature space by exploiting kernel-based dimensionality reduction methods. In this reduced subspace, they found sparse combination coefficients for a test sample and assigned a class label to it.

Support vector machines (SVMs), kernel Fisher discriminant analysis (KFDA) and kernel principal component analysis (KPCA) were examined for popular kernel-based learning methods in [2]. It discussed about the Vapnik–Chervonenkis (VC) theory and kernel feature spaces. They carry on to kernel based learning in supervised and unsupervised scenarios which includes the practical and algorithmic considerations. This paper illustrates the usefulness of kernel

algorithms with the applications such as optical character recognition (OCR) and DNA analysis.

According to [3] a new technique for direct visual matching of images for the purposes of face recognition and image retrieval, using a probabilistic measure of similarity, based primarily on a Bayesian (MAP) analysis of image differences is proposed. A comparison of probabilistic matching technique over standard Euclidean nearest-neighbour Eigen face matching was demonstrated. The performance results are compared using DARPA's 1996 FERET face recognition competition. In addition, it derived a simple method of replacing costly computation of nonlinear (on-line) Bayesian similarity measures by inexpensive linear (on-line) subspace projections and simple Euclidean norms, thus resulting in a significant computational speed-up for implementation with very large databases.

In [4] a direct method of calculation of weights is proposed by Fernández-Delgado et-al, also it gives an idea of weight estimation without the training stage of the classifier. Kernel trick is utilized to handle nonlinear data. Comparison of Gaussian kernel with SVM and LDA has been shown with the results of benchmark two-class dataset. Computation Cost gets linearly increased with the increase in the input dimension space.

In [5] the classification accuracy get very much increased and analysis of this accuracy has been analyzed with intrinsic parameters and with different kernel types. Normally multilayer perceptron maps a input feature vector into a higher dimensional space, but this paper has given an idea to project an intermediate Hilbert feature space.

3.KERNEL COLLOBORATON WITH PERCEPTRON

Suppose that we have n training samples for 'c' classes where the training samples are represented by $X=\{x_1, \dots, x_n\}$ ($j=1,2, \dots, n$) and p is the input dimension of each input sample. Given an arbitrary sample x, the goal is to assign a class label to it. The training samples are nonlinear in nature in the given input sample. In order to make it linear separable the data are mapped onto the higher dimensional kernel feature space. The nonlinear mapping function is carried out using any one of the kernel tricks. Eqn (2.1) shows the nonlinear mapping function corresponding to a kernel k(.,.).

$$\phi : \Phi(x) = [\phi_1(x), \phi_2(x), \dots, \phi_D(x)]^T \quad (2.1)$$

$D \gg m$ is the dimension of the feature space. Now the perceptron algorithm is applied for estimating the weights for all classes. The weights are re estimated until all the training samples are classified well.

$$X = \{x_1, x_2, \dots, x_n\} \quad Y = \{y_1, y_2, \dots, y_n\}$$

Weights are initialized to zero. From the training set the misclassified data are collected. The weights are reestimated according to the misclassified data.

3.1 Algorithm

Input: A set of training samples and their class labels

$$X = \{x_1, \dots, x_m\} \quad Y = \{y_1, \dots, y_m\}$$

where X denotes training samples and Y denotes the class labels for 'm' examples

- (i) Select a Mercer Kernel k(.,.), and its parameter.
- (ii) Calculate the kernel Gram matrix K where $K_j = k(x_i, x_j)$

3.1.1. Applying Perceptron

- (iii) Initialize $f(.) = 0$
- (iv) Repeat
 - (a) Pick (x_i, y_i) from data
 - (b) if $y_i f(x_i) \leq 0$ then (misclassified data)

$$f(.) = f(.) + y_i k(x_i, .) + y_i$$
 Until $y_i f(x_i) > 0$ for all i
 where $f(x) = \sum_j y_j k(x_j, x) + b$

Depending on the number of permutation between the classes, the weight vectors are generated. For a 2 class problem the number of weight vector generated is one as the number of permutation between the class remains one.

For a given test sample x, the product of weight vector and the sample will give a minimum error value. From this it is inferred that the test sample belongs to that class in which the weight vector belongs to.

The weight vectors generated during the training phase are utilized in the test phase for assigning the class labels of test sample.

3.2. Design of Training Phase

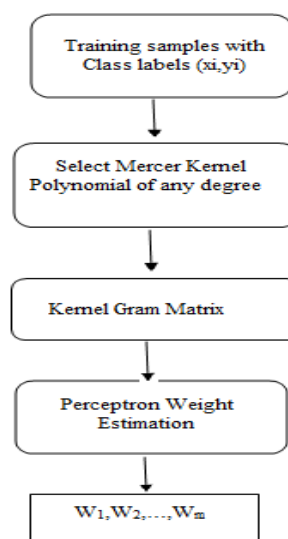


Fig 1 Training Phase

During the training phase, the training samples whose class labels are known are taken to train the classifier for the weight estimation. The samples are denoted by (x_i, y_i) where x_i and y_i denotes the training samples, Class labels for the sample i respectively. The training samples are projected in higher dimensional feature space using Kernel Trick. Mercer Kernel is used with any polynomial degree is utilized for generating the Kernel Gram Matrix. The weight vectors are first initialized to zero or to a random number. Then the

misclassified training samples are separated. Then the perceptron method is used for adjusting the weights between the misclassified training samples of different classes. The weight vectors are re-estimated until all the misclassified training samples are classified properly. Then after the reestimation iteration, the weight vectors are finalized. The number of weight vectors generated depends on the number of possibility of combination of the classes.

For a 2- class problem the number of weight vector is one(Combination: $W_1(\text{Class1,Class2})$).

For a 3- class problem the number of weight vectors generated are three(Combination: $W_1(\text{Class1,Class2}), W_2(\text{Class1,Class3}), W_3(\text{Class2,Class3})$).The weight vectors generated in the training samples are used in the testing phase for estimation of class labels of test samples.

3.2.Design of Testing Phase

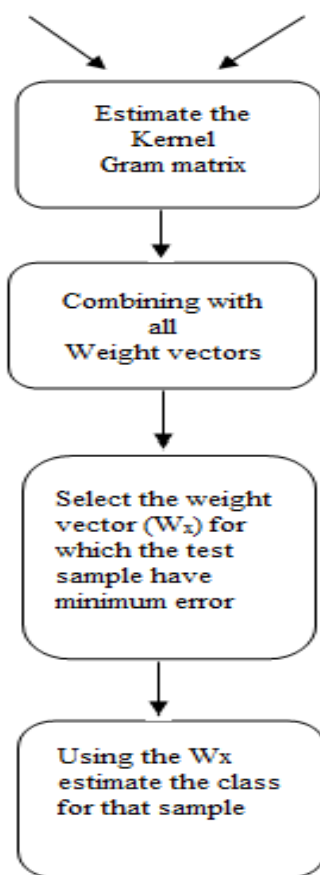


Fig 2 Testing Phase

The kernel gram matrix for the test sample is generated by combining with the training samples. Then the weight vectors are combined with the high dimensional test sample. Minimum error value will get generated for certain weight vectors for which the test sample belongs to. Then that weight vectors are selected for prediction of class labels. Then the selected weight vectors are used to estimate the class labels.

4.IMPLEMENTATION

The simulation of Kernel collaboration of perceptron for emotion recognition is done in Matlab [12]. In the simulation model, the classifier is designed with two class, multiclass nonlinear datasets. Table 1 shows the various parameters used for simulation.

Table 1.Implementation Parameters

Parameter	Value
Number of class	2
Number of examples	1500
Number of Features	2
Number of class 1 examples	300
Number of class 2 Examples	1200
Number of Training Samples	1125
Number of Test Samples	375

The classification for the training samples and test samples are carried out using the estimated weight vectors. Then the classified data are plotted to evaluate the accuracy efficiency.

Table 2.Implementation Results

Parameter	Expected Result	Estimated Result
Training Data-Class1	225	225
Training Data-Class2	900	900
Test Data-Class1	75	75
Test Data-Class2	300	300

The results shown in the fig 1 shows the training data classification over the decision surface.The decision surface as well as training data are classified according to the weight vectors estimated by the classifier. The results shows that the classification accuracy is 100% for training data.

The results shown in the fig 2 shows the test data classification over the decision surface. The decision surface as well as test data are classified according to the weight vectors estimated by the classifier. The results shows that the classification accuracy is 100% for test data.

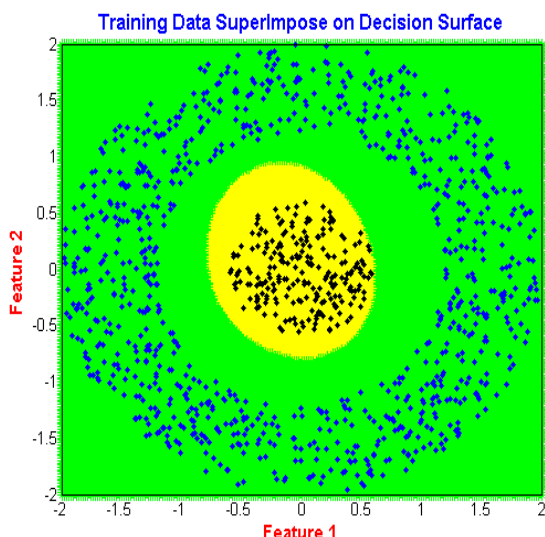


Fig 3.Training Data Superimpose on Decision Surface

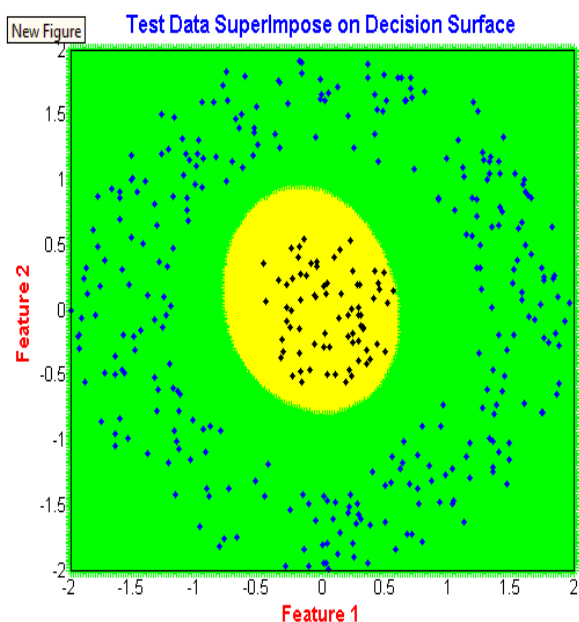


Fig 4.Test Data Superimpose on Decision Surface

4.1 Evaluation of Facial Dataset:



The classifier gets extended to recognize the emotion of a person based on the facial expressions. The dataset considered here is from JAFEE dataset which is available in [8].The original expression features are extracted from the image, where each column represents the features of a single image. Since the dataset is of TIFF type, it is of dimension 256 X 256. The total number of images for training and testing samples are 10. The above dataset has six classes which have been depicted in the above figure. The classes under consideration for the emotional recognition are angry, Distress, Fear, Happy, sad, surprise. The simulation results show 100% efficiency.

5. CONCLUSION

In this paper a classifier for handling multiclass nonlinear dataset is used for recognizing the emotion of a person. The classifier uses the kernel trick to make the non-linear data separable in a linear fashion. The classification is carried out using the perceptron technique. The weights are re-estimated until all the misclassified dataset are classified properly. The perceptron is trained with the kernel projection of the input data. Hence it is worth that the classifier is trained to handle the nonlinear data efficiently. The classifier is extended to handle multiclass problem. The classifier is tested using the toy dataset for two class and 4 class. Then the multiclass classifier is used to recognize the emotion of the person based on the facial expressions. The classifier produces 100% classification accuracy in working with toy dataset and JAFFE dataset.

6. REFERENCES

- [1] Li Zhang, Member, IEEE, Wei-Da Zhou, Member, IEEE, Pei-Chann Chang, Jing Liu, Member, IEEE, Zhe Yan, Ting Wang, and Fan-Zhang Li, "Kernel Sparse Representation-Based Classifier", in IEEE Transactions on Signal Processing, Vol. 60, No. 4, April 2012
- [2] Klaus-Robert Müller, Sebastian Mika, Gunnar Rätsch, Koji Tsuda, and Bernhard Schölkopf An Introduction to Kernel-Based Learning Algorithms IEEE Transactions on Neural Networks, Vol.12, No. 2, March 2001
- [3] BabackMoghaddam!,*, Tony Jebara", Alex Pentland"An efficient kernel discriminant analysis method " in Elsevier Pattern Recognition 33 (2000) 1771}1782
- [4] Fernández-Delgado, M.Santiago de Compostela,Ribeiro, Jorge;Cernadas, E.; Barro, S. "Fast weight calculation for kernel-based perceptron in two-class classification problems" in International Conference On Neural Networks (IJCNN 2010) ,18-23 July 2010
- [5] Rauber, Thomas Walter "Kernel Multilayer perceptron" in 24th International conference on Graphics, Patterns and Images (Sibgrapi 2011),28-31 Aug. 2011[6]
- [6] HichemSahbi, Jean-Yves Audibert, and Renaud Keriven" Context-Dependent Kernels for Object Classification", IEEE Transactions onPattern Analysis and machine Intelligence, Vol. 33, No. 4, April 2011
- [7] BabackMoghaddam!,*, Tony Jebara", Alex Pentland" Mitsubishi Electric Research Laboratory, 201 Broadway, 8th yoor, Cambridge, MA 02139, USA Massachusetts Institute of Technology, Cambridge, MA 02139, USA "Bayesian face recognition" in Elsevier Pattern Recognition 33 (2000) 1771}1782