OPTIMIZED CLASSIFIER DESIGN USING WAVELETS AND INDEPENDENT COMPONENT ANALYSIS FEATURE EXTRACTION FOR SUPPORT VECTOR MACHINES

SĂRARU, Corina
Department of Mathematics and Computer Science
University of Piteşti
corina.sararu@upit.ro

STATE, Luminiţa
Faculty of Mathematics and Computer Science
University of Piteşti
lstate@clicknet.ro

Abstract

The aim of the paper is to propose a new methodology for solving classification tasks based on feature selection using CWT (Complex Wavelet Transforms) and ICA (Independent Component Analysis) projections, combined with SVM (Support Vector Machines).

The resulting supervised classification ensemble system is designed based upon taking projections computed by CWT and estimating independent components from the data set. Each of the classifiers decides the assignation of the test data to certain classes using a different projection of the data. The ensemble classification involves combining the individual decisions of the particular considered classifiers.

The data that is fed to the classifiers is based on real and imaginary, magnitude and phase projections of the characteristics obtained in the feature selection step of the algorithm.

The proposed ensemble system of classifiers based on CWT and ICA supply higher global success rate as compared to the standard classifiers performance. Some of our results confirming this hypothesis are presented and commented in the final section of the paper.

Keywords: *Complex Wavelet Transform, Independent Components Analysis, Support Vector Machines, Pattern Recognition.*

ACM/AMS Classification: Primary 68T10, 62H30, 68U10; Secondary 68T05

1. **Introduction**

Feature selection represents an important step in elaborating a classification system. The optimal computation of features for the data depends on the particular classification problem that is considered.

In signal processing, feature extraction is done either by direct computation on the data, or using time-frequency analysis.

Some of the most popular instruments used to characterize the information present in the signals are: Principal Component Analysis, Independent Component Analysis, Fourier transform, wavelet transforms.

There are several approaches that have been explored regarding feature selection for Support Vector Machines (SVM) combined with selecting independent and/or principal components.

A new approach using a version of KPCA (Kernel Principal Component Analysis) of the problem of feature extraction from hyperspectral images is presented in [9]. The disadvantage of the PCA use is that it cannot characterize the data set in terms of higher order statistics [5]. The use of ICA-SVM combination was proposed in many pattern recognition problems such as face detection [11], or multivariate non-Gaussian process monitoring [6]. The simplest method for combining ICA and SVM is to estimate the independent components and use them as input data for the SVC in order to be classified ([11]).

Another approach of combining ICA and SVM in a pattern recognition problem was the use of a combination of features extracted through cyclic spectral analysis (using ICA) and the building of an ICA-SVM hybrid recognition system [3]. The ICA step was used to refine the features used in the pattern recognition.

Other approaches attempted to estimate the best features for SVM, in terms of extracting independent components and principal components ([7]). It has been show how iteratively estimating the features of the data set, gives a more efficient classification can be executed, because the separability of the data in terms of independent or principal components is higher ([7]).

2. **Feature Extraction Using CWT and ICA**

A wavelet transform may detect and emphasize the transition areas of a signal. The sudden transitions in a signal determine high amplitude for wavelet coefficients.

A wavelet family is defined based on a function ψ called mother wavelet:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{(s)}}\psi\left(\frac{t-u}{s}\right) \tag{1}$$

where, the parameters s and u are used respectively for scaling and translating the function ψ.

The transform

$$Wf(u,s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} \psi^* \left(\frac{t-u}{s}\right) f(t)dt = \langle \psi_{u,s}, f \rangle \tag{2}$$

is called the *wavelet transform* of the signal f defined by the mother wavelet ψ ([10]).

The study of complex wavelet transforms became relevant for feature extraction due to the latent information obtained by projecting the signals transformed in the two-dimensional space of the complex hiperplane [8].

ICA is relatively a model-free blind source separation method that attempts to separate all underlying sources contributing to the data without knowing these sources or the way that they are mixed:

$$s = Bx \tag{3}$$

In order to determine the matrix B, usually, an intuitively justified criterion function is selected, yielding to an unconstrained optimization problem.

It can be shown ([5]) that minimizing I corresponds to the maximization of the negentropy, a measure of non-Gaussianity of the components.

In our work, we used the FastICA algorithm introduced by [5]. In [5] the negentropy is approximated by

$$F(y) = [E\{G(y)\} - E\{G(\nu)\}]^2 \tag{4}$$

where $y = w^T x$, G is an nonquadratic function, ν is a Gaussian variable of zero mean and unit variance, yielding to the constraint optimization problem [5]:

$$maxF(w^T x), ||w||^2 = 1 \tag{5}$$

3. **Support Vector Machines**

Let $X = (X_1, ..., X_N)$ be a data set of n-dimensional training inputs $X_i \in \mathbf{R^n}, i \in \{1, ..., N\}$ where $Y = (y_1, ..., y_N)$, and $y_i = -1$ if X_i belongs to Class 1, or $y_i = 1$ if X_i belongs to Class 2. If these data are linearly separable, the study the behavior of a decision function is sufficient:

$$D(X_r) = w^T X_r + b, r \in \{1, ..., N\} \tag{6}$$

If $D(X_r) > 0$ then X_r is classified in Class 1, if $D(X_r) < 0$ then X_r is classified in Class 2. In order to find the optimal separating hyperplane represented by the decision function D, defined in Equation (6), the objective function Q is used [1]:

$$Q(w) = \frac{1}{2}||w||^2 \tag{7}$$

with w and b subject to the constraints:

$$y_i(w^t X_i + b) \geq 1, i \in \{1, ..., N\} \tag{8}$$

The problem of determining the optimal separating hyperplane reduces to minimization of Q. The problem can be defined as:

$$min(Q(w, b, \xi)) = min\left(\frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\xi_i\right) \tag{9}$$

subject to the constraints

$$y_i(w^t X_i + b) - 1 \geq 1 - \xi_i, i \in \{1, ..., N\} \tag{10}$$

where C is a margin parameter that balances the maximization of the margin and the minimization of the classification error.

The dual problem can be formulated as a maximization of:

$$Q(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j X_i^t X_j \tag{11}$$

subject to the constraints

$$\sum_{i=1}^{N}\alpha_i y_i = 0 \text{ and } C \geq \alpha_i \geq 0, i \in \{1, ..., N\} \tag{12}$$

The decision function is

$$D(X_r) = \sum_{i \in S}\alpha_i y_i X_i^t X_r + b, r \in \{1, ..., N\} \tag{13}$$

where S is the set of indices for the vectors X_i, for which $\alpha_i > 0$ (in this case, X_i is called support vector).

4. **The Proposed Algorithm**

In this section, we describe the proposed classifier design. Let X represent a data set, as described in the previous Section, where the samples are represented on rows. The classification process implies the use of the feature extraction techniques and the SVM classification methodology.

In the experiments, the function $feature\_extraction(.)$ is used, in order to represent the feature extraction technique (CWT and ICA):

$feature\_extraction(X)$

1. Preprocess X

2. ComputeX_W using the wavelet transform described by Equation 2

3. Compute

$$\begin{aligned} X_{ICA1} &= ICA(real(X_W)) \\ X_{ICA2} &= ICA(imaginary(X_W)) \\ X_{ICA3} &= ICA(magnitude(X_W)) \\ X_{ICA4} &= ICA(phase(X_W)) \end{aligned}$$

4. return $X_{ICA1}, X_{ICA2}, X_{ICA3}, X_{ICA4}$

The algorithm combines the use of individual SVM classifiers on each of the projections computed in the $feature\_extraction(X)$ function:

Algorithm 1:

1. Initialize Xtr (training data set), Xts (test data set), using the data set X;

2. $[Xtr_{ICA1}, Xtr_{ICA2}, Xtr_{ICA3}, Xtr_{ICA4}] = feature\_extraction(Xtr)$
   $[Xts_{ICA1}, Xts_{ICA2}, Xts_{ICA3}, Xts_{ICA4}] = feature\_extraction(Xts)$

3. Train four SVM classifiers using $Xtr_{ICA1}, Xtr_{ICA2}, Xtr_{ICA3}, Xtr_{ICA4}$;

4. Classify $Xts_{ICA1}, Xts_{ICA2}, Xts_{ICA3}, Xts_{ICA4}$ using the individual SVM classifiers trained at previous step;

5. Combine individual decisions to yield the ensemble decision of the classification system.

5. **Experimental Results**

Portions of the research in this paper use the MIT-BIH database of electrocardiograms (ECGs) that can be found at
http : //www.physionet.org/physiobank/database/.

The electrocardiogram (ECG) signal is the electrical interpretation of the heart activity; it consists of a set of well defined, successive waves denoted: P, Q, R, S, and T waves. The denoising based on wavelet theory [2] has been extensively exploited in filtering noisy ECG.

The ECG signals were taken from patients with/without supraventricular arrhythmia, for each sample being provided the correct diagnosis.

In our tests, we considered 20 records from which 10 records came from patients with supraventricular arrhythmia. Each record represents the signal measured using a 128Hz sampling frequency for 10 seconds. For training and testing, 2-fold cross validation was used on the 20 records.

The ICA-SVM classifier uses two types of kernels, polynomial of order 9, RBF and MLP) and the classification is computed on the coordinates of the data sets Xtr and respectively Xts in the independent component space.

The classification computed by Ensemble+ICA results by combining the individual decisions, using the logistic function and the threshold $\theta = 0.5$.

Some of the results of our tests on MIT-BIH database are summarized in Table 1.

Type of Classifier	Polynomial	RBF	MLP
ICA	0.55	0.50	0.55
Real+ICA	0.58	0.65	0.55
Imaginary+ICA	0.63	0.55	0.55
Magnitude+ICA	0.85	0.76	0.85
Phase+ICA	0.33	0.35	0.43
Ensemble+ICA	0.88	0.85	0.80

Table 1: Success Rate for Several Types of Classifiers

While the performance of the individual classifiers, using projections of ICA on the real, imaginary, phase or angle space, is very poor, the success rate of the proposed classifier design proves the efficiency of the use of ensemble decision making, by combining the individual decisions of each projected oriented classifier.

The Ensemble+ICA uses voting procedures for classification purposes, each of the voting classifiers being of the SVM type but applied to different sets of features extracted from the input signals.

The tests performed on the MIT-BIH database point out significant improvements in case of the proposed ensemble type classifiers and encourage further work.

References

1. S. Abe, *Support Vector Machines for Pattern Classification*, 2005

2. A. Boroomand, M. A. Ahmadian, A. Oghabian, J. Alirezaie, and C. Beckman, *An efficient hybrid wavelet-ica algorithm for analyzing simulated fmri data in noisy environment*, The 7th IEEE International Symposium on Signal Processing and Information Technology, 2007.

3. D., Boutte, B., Santhanam, *A Hybrid ICA-SVM Approach to Continuous Phase Modulation Recognition*, IEEE Signal Processing Letters, 16(5), 2009.

4. G.Y. Chen, G. Dudek, *Auto-correlation wavelet support vector machine*, Image and Vision Computing, Elsveier, 2009.

5. A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, 2001.

6. C., C., Hsu, M.C., Chen, L.S., Chen, *Intelligent ICA-SVM fault detector for non-Gaussian multivariate process monitoring*, Expert Systems with Applications, 2010.

7. J. Fortuna, D. Capson, *Improved support vector classification using PCA and ICA feature space modification*, 2004

8. N., G., Kingsbury, *Complex wavelets for shift invariant analysis and filtering of signals*, Journal of Applied and Computational Harmonic Analysis, 10(3), pp. 234–253, 2001.

9. W., Liao, A., Pizurica, W., Philips, Y., Pi, *A FAST ITERATIVE KERNEL PCA FEATURE EXTRACTION FOR HYPERSPECTRAL IMAGES*,Proceedings of IEEE 17th International Conference on Image Processing, 2010.

10. S., Mallat, *A Wavelet Tour of Signal Processing, 3rd Edition, Academic Press*, New York, 2009.

11. Y. Qi, D. Doermann, D. DeMethon, *Hybrid independent component analysis and support vector machinelearning scheme for face detection*, IEEE Int. Conf. Acoustics Speech and Signal Processing, 2, pp. 1481–1484,2001.