A New Heuristic Classifier Design Based on Independent Component Analysis and Wavelet Methods

Corina Săraru, Lumița State, Maria Miroiu
Faculty of Mathematics and Computer Science
University of Pitești
Pitești, Romania
corina_sararu@yahoo.com, lstate@clicknet.ro, miroiu_m@yahoo.com

Abstract

The aim of the paper is to propose a new methodology for solving classification tasks based on ICAW (Independent Component Analysis and Wavelets).

The general idea is to use an ensemble system of classifiers that decide independently on representations resulted by taking projections computed by wavelets and Independent Component Analysis (ICA) and to combine the decisions of the particular classifiers in order to improve the overall decision.

Each classifier establishes its decision on a different projection of the data and for each individual combined with an ensemble system of classifiers, the best classification for any individual is computed by combining the decisions of the particular considered classifiers. Each classifier uses a different projection of the data.

The individual classifiers perform classification on real and imaginary coefficients, magnitude and phase corresponding to the representations of the data. The ensemble system of classifiers is essentially based on ICA projections supply higher global success rate as compared to the individual classifiers performance. Some of our results confirming this hypothesis are presented and commented in the final section of the paper.

Keywords: independent components analysis, wavelet decomposition, pattern recognition, signal processing.

1. Introduction

Independent components analysis (ICA) has been emerging as an efficient tool to isolate the independent components in signals. It does not need any prior knowledge about each source, only utilizes the general information that all the sources are independent with each other. In this application, the number of the sensors should match the number of the sources to acquire enough information to support ICA work.

There is a basic assumption in this method: the signal from each sensor has different mixing ratio of the independent components from those from others. ICA has been introduced to mechanical dynamic signal analysis in the last few years [7],[12].

Other techniques have been developed in order to detect abnormalities in informational data. One of them is Principal Component Analysis PCA. Both PCA and ICA are used to identify certain components existing in the signal. They are based, though, on different rules. For PCA, the principal component should follow the direction that has the maximum data variance. While for ICA, each of the components is extracted such that they are independent with one another.

ICA is more suitable when the purpose is to find a component from a mixture of many independent sources. However, in some circumstances, there is limitation to install too many sensors to satisfy the requisition for ICA. For example, to diagnose a small gearbox, there are some bearings and gears inside and two couplers are connected between the output shaft and other equipment.

One of the important issues when using ICA is the noise present in the observational data. The noise reduction can be done based on two different approaches:

1. Construction of a theoretical model that includes the notion of noise and eliminates it through specific techniques [9].

2. Preprocessing the observed data such that most of the noise is removed.
The main disadvantage of the first type of approach is that it is time consuming and computationally expensive. That is why the idea of preprocessing the data is much more popular nowadays. The noise reduction may be done by applying suitable low pass filters in time domain before implementing ICA algorithm [2].

A wavelet transform can focus on localized signal structures with a zooming procedure that progressively reduces the scale parameter. Singularities and irregular structures often carry essential information in a signal. For example, discontinuities in images may correspond to occlusion contours of objects in a scene.

A wavelet transform of the observed data may be regarded as a filtering procedure. It is performed on localized signal parts, a scaling procedure allowing to change the values of the corresponding parameters. It is generally accepted that singularities and irregular parts of an signal often contain essential information about the objects in the signal.

Singularities and edges are detected from wavelet transform local maxima at multiple scales. These maxima define a geometric scale/space support from which signal and image approximations are recovered. Non-isolated singularities appear in highly irregular signals such as multi-fractals.

The idea behind applying wavelet decomposition before feeding the observation data to ICA is to improve the assumption of non-gaussianity distribution of sources enforced for ICA algorithm and increasing the independency of sources. The projection of data to a set of orthogonal basis function in wavelet domain produces fewer coefficients to represent the data leading to super-gaussian distribution of data.

Removing noise from signals is possible only if some prior information is available. This information is encapsulated in an operator designed to reduce the noise while preserving the signal. Ideally, the joint probability distribution of the signal and the noise is known. Bayesian calculations then derive optimal operators that minimize the average estimation error. However, such probabilistic models are often not available for complex signals such as natural images.

Singularities and edges are detected using the local maxima of wavelet transform at different scales. These maxima define a geometric scalespace support from which signal or image approximations are recovered [15].

In order to improve the accuracy with noise present in data, the k-NN algorithm introduces a parameter k so that for each new example q to be classified the classes of the k nearest neighbors of q are considered: q will be labeled with the majority class. Another alternative consists in assigning that class whose average distance is the smallest one or introducing a heuristically obtained threshold $k 1 < k$ so that the assigned class will be that with a number of associated examples greater than this threshold [20].

A more sophisticated approach, k-nearest neighbor (k-NN) classification [5], finds a group of k objects in the training set that are closest to the test object, and bases the assignment of a label on the predominance of a particular class in this neighborhood. This addresses the issue that, in many data sets, it is unlikely that one object will exactly match another, as well as the fact that conflicting information about the class of an object may be provided by the objects closest to it. There are several key elements of this approach: (i) the set of labeled objects to be used for evaluating a test objects class (ii) a distance or similarity metric that can be used to compute the closeness of objects, (iii) the value of k, the number of nearest neighbors, and (iv) the method used to determine the class of the target object based on the classes and distances of the k nearest neighbors. In its simplest form, k-NN can involve assigning an object the class of its nearest neighbor or of the majority of its nearest neighbors.

In this paper an heuristic classifier design based on ICA and wavelets (ICAW) methods is presented.

Being given a database containing a set of observations coming from two classes such that the true provenance class is known for each individual, new representations are computed in terms of the features extracted using the wavelets described in Equation 6.

The FastICA algorithm is applied to the initial database and to each collection of representations in terms of the above mentioned features.

The design of the new classifier corresponds to a weighted voting procedure that combines the decisions of the ensemble of resulted classifiers where the weights are given by the corresponding correctness scores computed for each classifier.

### 2. Theoretical Framework: ICA and Wavelet Decomposition

Briefly, the principle of noise free ICA algorithm can be stated as follows:

Let $X$ be the observation matrix assumed as being linear combinations of some independent hidden sources,

$$X = AS$$

(1)

This expression leads to

$$S = WX$$

(2)

The goal of ICA is to discover the hidden set of sources using also linear combinations of the observational vectors,
that is to find the matrix \( W \) such that the rows of \( S \) are "as independent as possible". In this case, the rows of \( S \) are taken as estimations of the hidden unknown independent sources.

There have been proposed a series of ICA type algorithms [9], some of the most frequently used being FastICA and InfoMax [2].

The Fast ICA algorithm is based on maximizing non-Gaussian property of the estimated sources where the non-Gaussian is expressed in terms of the negentropy. Assuming that the mean \( \mu \) and the variance \( \sigma^2 \) of the random variable \( Y \) that models the observational environment, the negentropy is expressed by the difference between the entropy of the Gaussian variable \( N(\mu, \sigma^2) \) and the entropy of the random variable \( Y \) that models the observational environment.

That is [9],

\[
J = \ln(\sqrt{2\pi\sigma^2}) + 1/2
\]

According to the InfoMax principle, \( J \geq 0 \) and \( J = 0 \) if and only if \( Y \sim N(\mu, \sigma^2) \) [5].

Being given that only in very particular cases, the evaluation of the entropy \( H(y) \) can be effectively carried out, some approximation of \( J \) is needed.

In our work, we used the FastICA algorithm introduced in [9], where the negentropy is approximated by

\[
F(Y) = [E(G(Y)) - E(G(\nu))]^2
\]

where \( G \) is an nonquadratic function, \( \nu \) is additive white noise, \( \nu \sim N(0, 1) \) and \( Y = WX \).

Therefore, the problem of finding approximates of the hidden independent sources is reduced to the constraint optimization problem,

\[
\max \{ F(WX) \} \quad \text{subject to } ||W|| = 1
\]

A wavelet is a real valued function \( \psi \), having the property that

\[
\int_{-\infty}^{+\infty} \psi(t)dt = 0
\]

Using a given wavelet (referred as the mother wavelet) \( \psi \) to model the window shape and scaling and translation transforms, we get the family of functions, ([15])

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

where the parameters \( s \) and \( u \) are used for scaling and translating the windows.

So far there have been introduced a long series of types of wavelets, the option about the shape of the window being imposed by the particular problem that is handled.

We intend to use wavelet based techniques in processing sets of observations taken on electrocardiographic data. Being given the specificity of this problem, our option is to use as the mother wavelet \( \psi \) a member of the complex Gauss wavelet family, defined by:

\[
\psi(n) = f^{(n)}
\]

where \( f^{(n)} \) is the \( n \)-th derivative of the function \( f(x) = C_ne^{-ix^2}e^{-x^2} \) and \( C_n \) is chosen such that \( ||f^{(n)}||^2 = 1 \).

The features extracted from signals result by convoluting the signal \( f \) with the windows defined by the wavelets derived from the mother wavelet (6) ([14]):

\[
Wf(u, s) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt(s)} \psi \left( \frac{t - u}{s} \right) dt.
\]

3. The ICAW-k-NN Algorithm

The scheme of the classification system is:

![Figure 1. The classifier design](image)

The preprocessing module consists of applying denoising methods to clean the initial input data.

In the module Complex Wavelet Decomposition, the cleaned data are processed in order to extract significant features using (7).

The FastICA algorithm is applied in order to identify its latent independent structure.

Each of the resulted databases is classified using the kNN method and the performance is evaluated in terms of the empirical error against the true known classifications.
The resulted correctness scores are fed in the Weight Estimation module and combined in order to determine a set of classification weights used in the voting procedure that combines the classification decisions corresponding to each classifier into a final decision.

The performance of the resulted ensemble classifier is evaluated in terms of the empirical error against the true known classifications.

In case the classification error is not acceptable the system can ask for further data and/or to initiate more refined denoising techniques.

The classification task is solved using two classifiers referred in Table 1 as Ensemble1+ICA and Ensemble2+ICA. Both classifiers are essentially modified versions of the 2-NN by including a voting mechanism that involves the decisions of individual classifiers. In Ensemble1+ICA the voting mechanism is a simple majority rule.

The Ensemble2+ICA classifier uses a more refined classification rule of connectionist type. The decisions of individual classifiers are combined additively using the weights computed in the Weight Estimation module. The logistic function is applied to the resulted value and the class is computed by applying the threshold \( \theta \).

In our tests the data contains equal size samples coming from the normal/abnormal classes and for this reason we set \( \theta = 0.5 \).

In cases when the data contains \( N = N' + N'' \) samples, where \( N' \) and \( N'' \) are the sizes of data coming from the class labeled \( -1/1 \) and \( N' \neq N'' \), the value of the threshold is set by taking into account the sizes of data coming from each class in the design and test phase respectively.

The training phase aims to obtain the weights used to classify new data. We use \( N_1 \) samples from the total amount of \( N \) to estimate the weights and \( N_2 \) samples to evaluate the performance.

The \( N_1 \) samples are submitted to be classified by eight classifiers, (see Table 1) and the empirical errors are computed for each classifier.

In the Weight Estimation module the empirical errors are used to compute the weight of each classifier, the weight being the ratio of the success rate and the overall sum of success rates.

The testing procedure consists of submitting the remaining \( N_2 \) samples for being classified by Ensemble1+ICA and Ensemble2+ICA.

The results of our tests on the MIT-BIH database are presented in Table 1.

Based on the previously described theoretical concepts, the electrocardiogram (ECG) signal is a repeating and almost periodic pattern. This characteristic of physiological signals is extensively exploited in filtering noisy ECG. [3].

The analysis of the different DWT levels shows that the 1st level detail sequence of the noisy ECG signal is highly dominated by the wavelet gaussian noise (WGN) energy [3].

When the biomedical signals are corrupted by some artifact, a preprocessing step is needed in order to extract relevant clinical information from the data.

For this reason the artifact cancelation is a key topic in biomedical data processing [18]. In particular, the artifact removal is often necessary for the clinical study of the electrocardiographic (ECG) signal [18].

The ECG signal looks like a repeating and almost periodic pattern. This characteristic of physiological signals was explored in order to synchronize the parameters of the filter with the period of the signal.

For instance, Liang and Lin [13] proposed an efficient method based on discrete wavelet transform (DWT) in order to perform the cancelation of stimulus artifact in the serosal recordings of gastric myoelectric activity, but it works well only when there is no interference between the filter and the ECG signal.
Aiming to remove the artifacts in biomedical signals, even in the presence of interference with the ECG signals, several methods based on Independent Component Analysis (ICA) have been also proposed [1], [10].

The methodology that was previously described in Section 3 was applied on the MIT-BIH database, http://www.physionet.org/physiobank/database/, containing samples of ECG signals taken from patients with/without supraventricular arrhythmia, for each sample being provided the correct diagnosis.

In our tests, we considered 35 records from which 25 records came from patients with supraventricular arrhythmia.

Each record represents the signal measured using a 128Hz sampling frequency for 10 seconds.

The training data consisted of all records coming from patients without supraventricular arrhythmia and 10 records selected from the remaining 25 signals.

The classifier Std implements the standard 2-NN classification rule on the unprocessed samples, while the ICA classifier uses the same classification rule applied to the representations computed by the FastICA algorithm. The Real+ICA, Imag+ICA, Magn+ICA, Phase+ICA perform the classifications on the representations in terms of the features extracted by combined wavelet and ICA method, real(R), imag(R), abs(R), phase(R), respectively.

Ensemble1+ICA performs the classification by a majority voting procedure using the decisions taken by the classifiers Real+ICA, Imag+ICA, Magn+ICA, Phase+ICA.

The classification computed by Ensemble2+ICA results by combining the decisions of Real+ICA, Imag+ICA, Magn+ICA, Phase+ICA, using the logistic function and the threshold $\theta = 0.5$.

The results of our tests on MIT-BIH database are summarized in Table 1 and Figure 2.

The tests performed on the MIT-BIH database point out significant improvements in case of the proposed ensemble type classifiers and encourage further work on one hand, in refining the classification scheme and on the other hand, in identifying classes of wavelets better adapted to this particular problem.

5. Conclusive remarks and suggestions for further work

The research aimed to propose a new classification technique based on wavelet and ICA methods.

The novelty is represented by the Ensemble1+ICA and the Ensemble2+ICA that use voting procedures for classification purposes, each of the voting classifiers being of the 2-NN type but applied to different sets of features extracted from the input signals.

<table>
<thead>
<tr>
<th>Type of Classifier</th>
<th>Success Rate(Percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std</td>
<td>26.67</td>
</tr>
<tr>
<td>ICA</td>
<td>53.33</td>
</tr>
<tr>
<td>Real+ICA</td>
<td>73.33</td>
</tr>
<tr>
<td>Imag+ICA</td>
<td>66.67</td>
</tr>
<tr>
<td>Magn+ICA</td>
<td>46.67</td>
</tr>
<tr>
<td>Phase+ICA</td>
<td>66.67</td>
</tr>
<tr>
<td>Ensemble1+ICA</td>
<td>80.00</td>
</tr>
<tr>
<td>Ensemble2+ICA</td>
<td>86.67</td>
</tr>
</tbody>
</table>

Table 1. Success Rate for Several Types of Classifiers

The tests performed on the MIT-BIH database point out significant improvements in case of the proposed ensemble type classifiers and encourage further work on one hand, in refining the classification scheme and on the other hand, in identifying classes of wavelets better adapted to this particular problem.

References


