CLASSIFICATION OF TIME SERIES VIA WAVELET SUBBAND ANALYSIS USING SUPPORT VECTOR MACHINE CLASSIFIER

Béla PALÁNCZ* and Balázs BENYÓ**

*Department of Photogrammetry and Geoinformatics
Budapest University of Technology and Economics,
H–1111 Budapest, Hungary, Műegyetem rkp. 3.
e-mail: palancz@epito.bme.hu
**Department of Informatics
Széchenyi István University
H-9026 Győr, Hungary, Egyetem tér 1.
e-mail: benyo@sze.hu

Abstract

An improved feature extraction method has been developed for classification and identification of time series, in case of the number of the experiments are considerably less than that of the samples in time series. The method based on the subband analysis of the wavelet transformation of the time signals, provides lower dimension feature vectors as well as much more robust kernel-based classifier than the traditional wavelet-based feature extraction method does. The application of this technique is illustrated by the classification of cerebral blood flow oscillation using support vector classifier with Gaussian kernel. The computations were carried out with Mathematica 5.1 and its Wavelet Application.

Keywords: Support Vector Machine, wavelet subband analysis, classification of time series, cerebral blood flow oscillation.

1. Introduction

Sometimes the numbers of physiological experiments are restricted by many reasons (i.e. technical, financial, moral etc.). It can happen, that the number of the different experiments (m) is considerably less than the number of samples measured in each experiments (n), namely $n >> m$. This fact can make it more difficult to classify or identify such a time series. This study deals with the classification of cerebral blood flow signals (CBF), where we have faced the above problem.

Oscillation of the CBF is a common feature in several physiological or pathophysiological states and may significantly influence the metabolic state of the brain. Medical experiments were carried out to study the effect of special drugs on CBF in adult male Wistar rats' brain, [1]. A typical part of such a time signal is shown in *Fig. 1*.

```
ts=ReadList["E:\data1.txt",Number];
```

```
ListPlot[Drop[ts,70000],Frame → True,PlotJoined → True,

    Axes → None,FrameLabel → {"Samples","Blood Folw
AU"},AspectRatio → 0.4];
```

The total length of the signal, the number of the samples is
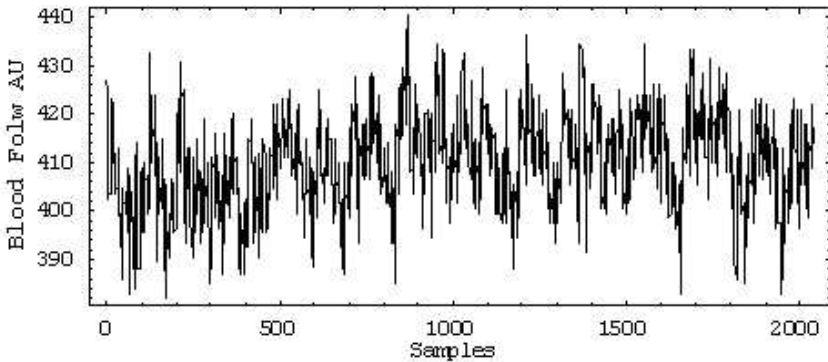
```
n=Length[ts]
    72041
```


Fig. 1. A part of a time signal of CBF with sampling time 5 ms.

In order to identify different states of CBF oscillation, different classification methods, based on a two-dimensional feature vector – the maximum amplitude and its frequency of the Fourier transform of the time signals – have been employed, using neural network and support vector machine classifiers (SVMC) [2] and [3]. However, these approaches were only partly successful because the two-dimensional feature vector could not characterize all the features of the time series. Even the most promising technique, the SVMC suffered from overlearning [4].

In this paper an improved feature extraction method has been developed for classification and identification of time series, when the number of the experiments is considerably less than that of the samples in the time series. The method is based on the subband analysis of the wavelet transformation of the time signals providing lower dimension feature vectors as well as much more robust kernel-based classifier than the traditional wavelet-based feature extraction method. The application of this technique is illustrated by the classification of CBF oscillation using support vector classifier with Gaussian kernel. The type of this presentation is a notebook form of *Mathematica*, because the computations were carried out with *Mathematica* 5.1 and its Wavelet Application. In this way, we hope, that nothing is left out or hidden and the method is much more understandable and its results are easily reproducible for anyone.

2. Feature Extraction Via Wavelet Transformation

In recent years, feature extraction methods were developed based on wavelet trans-
formation to recognize acoustic signals. They are applicable to the recognition of
ships from sonar signatures, cardiopulmonary diagnostics from heart sounds, safety
warnings and noise suppression in factories, and recognition of different types of
bearing faults in the wheels of railroad cars and so on [5]. Let us illustrate this
classical technique applying it to a CBF signal.

First, we drop the beginning and the end of this raw signal, getting a signal
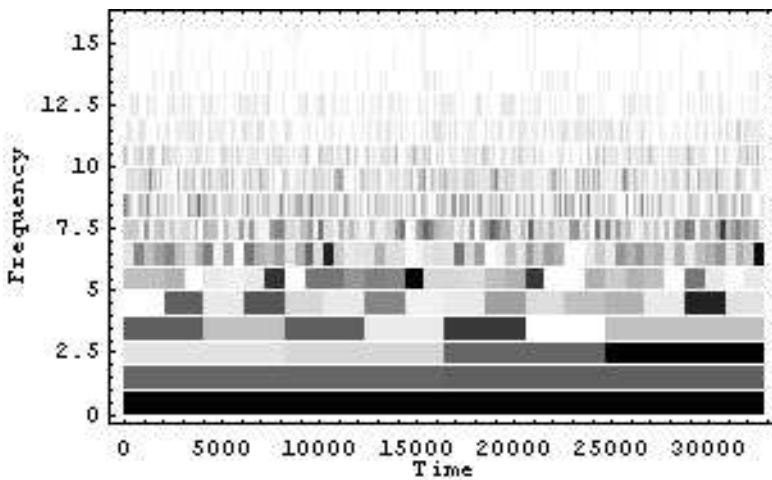of length of 2^{16} samples,

Fig. 2. The phase space plot of the DWT of the time signal

```
tsm=Drop[Drop[ts, -3000], n-3000-2^16]; length[tsm]
    65536
```

Loading Wavelet application,

```
<<Wavelets`Wavelets`
```

Now the DWT of the time signal can be computed,

```
ws=WaveletTransform[tsm,DaubechiesFilter[2]];
```

This transformation decomposes the data into a set of coefficients in the
wavelet basis. There are 16 sublists containing the wavelet coefficients in the
orthogonal basis of the orthogonal subspaces.

The contributions of the coefficients to the signal at different scales are rep-
resented by the phase space plot, see *Fig. 2*. Each rectangle is shaded according

to the value of the corresponding coefficient: the bigger the absolute value of the coefficient, the darker the area. The time unit is 5 msec.

```
PhaseSpacePlot[ws,Frame → True,LogarithmicScale → True,
FrameLabel → {"Time","Frequency"}];
```

The traditional feature extraction method, originally developed for recognition of acoustic signals of different types of bearing faults in the wheels of railroad cars, is the following [6]:

From the wavelet coefficients of each of the 16 resolution levels (subbands) and from sample values of the original time signal let us, compute the average energy content of the coefficients at each resolution. There are a total of 17 subbands (16 wavelet subbands and one approximation subband represented by the original signal) from which features are extracted. The ith element of the feature vector is given by

$$v_i = \frac{1}{n_i} \sum_{j=1}^{n_i} w_{i,j}^2, i = 1, 2, \ldots, 17 \tag{1}$$

where $n_1 = 2, n_2 = 2, n_3 = 2^2, \ldots, n_{16} = 2^{15}$ and $n_{17} = 2^{16}$, $w_{i,j}$ is the jth coefficient of the ith subband. In this way, from a time signal having 2^k samples or dimensions, one can extract a feature vector of $k + 1$ dimensions.

This technique has been extended for two dimensional signals, for digital images [7].

3. Employing Subband Analysis

In order to study the effect of the dimension of the input space on the quality of the classification as well as to save the morphology of DWT, here we employ a different approach. We consider the wavelet coefficients belonging to a given subband as a feature vector based on this given resolution. It can be a reasonable approach, because the approximated signal representation in the orthogonal subspace corresponding to this subband is given by these coefficients [8].

In our case, there are two sets of time signals, representing two classes of CBF states and only 40 patterns (2 × 20) are at our disposal. Intuitively, it is possible to shatter two points by any linear manner in the one-dimensional space and three points in two-dimensional space. By analogy, it is possible to shatter N+1 points in the N-dimesional space with probability 1. If the patterns to be classified are independent and identically distributed, then the 2 N patterns are linearly separable in the N-dimensional space [9].

The coefficients of the subbands from $n_2 = 2$ up to $n_6 = 2^5 = 32$ will be employed as different feature vector components. The magnitudes of the wavelet coefficients at these subbands are shown in *Fig. 3*.

```
PlotCoefficients[Drop[Drop[ws,1],-10],Frame → True,
FrameLabel → {"Wavelet coefficients","Resolution (subbands)"}];
```
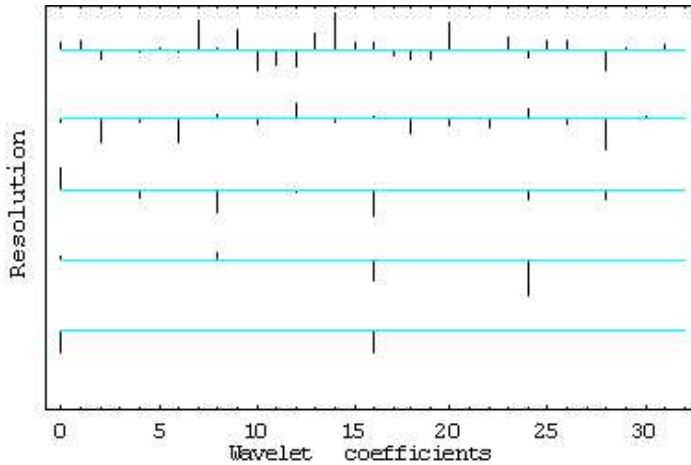

Fig. 3. The magnitude of the wavelet coefficients at resolution from $n_2 = 2$ (at the bottom) up to $n_6 = 2^5$(at the top)

It means that the very small coefficients belonging to the higher resolutions, $(n_7 - n_{16})$ and the very big coefficients of the lowest resolution, (n_1) are not taken into consideration. The previous ones have no contributions; the latest one would suppress all the others (see *Fig. 2*). With other words, we consider the "measurable" fine structure of the subband coefficients.

```
maxcoeffs=Table[{i,Max[Abs[ws[[i]]]]},{i,2,16}];
ListPlot[maxcoeffs,Frame → True,PlotRange → {{1,17},{0,320}},
PlotStyle → AbsolutePointSize[1],
FrameLabel → {"Resolution","Max of coefficients"},
Prolog → Map[Line[{{#[[1]],0},#}]&,maxcoeffs]];
```

Fig. 4 shows the maximums of the magnitude of the wavelet coefficients of different resolutions, except of those belonging to the first (lowest) one. To illustrate the overwhelming magnitudes of the first resolution, here are the coefficients of the first and second levels,

```
Join[ws[[1]],ws[[2]]]
```
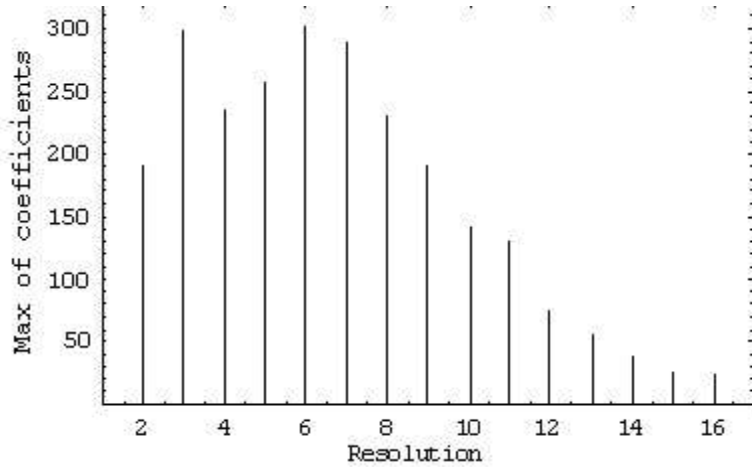$\{74276.2, 74516.4, -171.985, -189.396\}$

Fig. 4. The maximal magnitudes of the wavelet coefficients of different resolutions

4. SVM Classifier

For the classification, a support vector machine (SVM) classifier is used. This kernel-based classifier can be trained on any size of training set, while neural networks should have so many input nodes as the dimension of the input space and need definitely more training patterns than the number of these input nodes. Employing kernels, a classification problem can be transferred in a higher dimensional space, where the linear separability is more likely. In addition, the quality of the classification in any dimension can be measured by the geometric margin of the SVM classifier [10].

As an example let us load the coefficients of the fifth subband, $n_5 = 2^4 = 16$, for all the 40 patterns, and we get

```
wsp=ReadList["F:\data2.txt"];
```

so we have 40 feature vectors of dimension of 16, half of these feature vectors belonging to the first, the other half belonging to the second CBF state,

```
Dimensions[wsp]
    {40, 16}
```

First, these data should be standardized; to be transformed so that their mean is zero and their unbiased estimate of variance is unity,

```
<<Statistics`MultiDescriptiveStatistics`
wsps=Standardize[wsp];
```

Let us employ Gaussian kernel, with parameter $\beta = 5$,

```
β = 5.;
K[u_,v_]:=Exp[-β(u-v).(u-v)]
```

It is useful to compute the determinant and the condition number of the of the Gram matrix, [10]

```
KK=Table[K[wsps[[i]],wsps[[j]]],{i,1,40},{j,1,40}];
<< LinearAlgebra`MatrixManipulation`
```

The value of the determinant

```
Det[KK]
    1.
```

The condition number using infinity-norm is,

```
MatrixConditionNumber[KK]
    1.
```

The labels are indicating the first and second set, $y \in -1, 1$, therefore

```
zm=Join[Table[1,{i,1,20}],Table[-1,{i,1,20}]]
```

$\{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,$
$-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,$
$-1, -1, -1, -1, -1\}$

Let the value for the control parameter of regularization be

```
c = 100.;
```

To carry out the training of the support vector classifier, we shall employ the algorithm imbedded into the function, `SupportVectorClassifier` developed for *Mathematica* [11].

```
F1=SupportVectorClassifier[wsps,zm,K,c];
```

The short, analytical form of the classifier is

```
Short[F1[[1]],5]
```
$$-1.63702 \times 10^{-13} + << 59 >> +0.990099$$
$$e^{-5 \cdot ((-0.270843+x_1)^2+(-0.056462+x_2)^2+<<12>>+(-0.124474+x_{15})^2)+(3.35514+x_16)^2)}$$

Let us check the result of the classification. The values of continuous classifier are

```
Map[F1]]/.Table[xᵢ → #[[i]], {i,1,16}]& wsps]
```
$\{0.990099, 0.990099, 0.990099, 0.990099, 0.990099, 0.990099,$
$0.990099, 0.990099, 0.990099, 0.990099, 0.990099, 0.990099, 0.990099,$
$0.990099, 0.990099, 0.990099, 0.990099, 0.990099, -0.990099, -0.990099,$

− 0.990099, −0.990099, −0.990099, −0.990099, −0.990099, −0.990099,
− 0.990099, −0.990099, −0.990099, −0.990099, −0.990099, −0.990099,
0.990099, 0.990099, −0.990099, −0.990099, −0.990099, −0.990099,
− 0.990099, −0.990099}

Employing the signum decision function, we get

```
Partition[Sign[%],20]//MatrixForm
```

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1) \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix}$$

The classification is correct. The values of the weighting coefficients of the continuous classifier:

```
alfa=F1[[2]]
```

{0.990099, 0.990099, 0.990099, 0.990099, 0.990099, 0.990099,
0.990099, 0.990099, 0.990099, 0.990099, 0.990099, 0.990099,
0.990099, 0.990099, 0.990099, 0.990099, 0.990099, 0.990099,
0.990099, 0.990099, 0.990099, 0.990099, 0.990099, 0.990099,
0.990099, 0.990099, 0.990099, 0.990099, 0.990099, 0.990099,
0.990099, 0.990099, 0.990099, 0.990099, 0.990099, 0.990099,
0.990099, 0.990099, 0.990099, 0.990099}

what means, that all patterns are support vector. A sample pattern was considered as support vector, if its contribution (its weighting coefficient) to the decision function is greater than 1% of the maximal contribution.

The geometric margin, γ can indicate the quality of the classification [10], greater the γ, more reliable the classification:

$$\gamma = \left(\sum_{i=1}^{40} \texttt{alfa[[i]]} - \frac{1}{c}\texttt{alfa} \cdot \texttt{alfa} \right)^{-1/2}$$

0.159695

These computations were carried out for different feature vectors based on the coefficients of the different subbands *Table 1* shows the results.

Decreasing the number of the wavelet coefficients, the Gram matrix is getting ill-conditioned, the geometric margin is becoming narrower and probability of the misclassification of patterns is increasing, although the classification with four wavelet coefficients is just acceptable.

Let us employ the traditional feature extraction method, when the elements of the feature vector are computed as the average of squares of the wavelet coefficients belonging to the same subband, plus the same contribution of the original signal as additional "subband". Consequently, the dimension of the feature vector is 16+1=17. *Table 2* shows the result for this case.

Table 1. The results of the SVM classification with different feature vectors

Subband level	Number of wavelet coefficients	Determinant of Gram matrix	Condition number of Gram matrix	Number of support vectors	Geometric margin	Number of misclassified patterns
6	32	1.	1.	40	0.159695	0
5	16	1.	1.	40	0.159695	0
4	8	0.999	1.040	40	0.159701	0
3	4	0.005	69.374	40	0.113922	0
2	2	1.9310^{-39}	1.1510^{7}	25	0.083355	4

Table 2. The results of the SVM classification employing traditional feature extraction technique

Determinant of Gram matrix	Condition number of Gram matrix	Number of support vectors	Geometric margin	Number of misclassified patterns
0.994	1.170	40	0.159374	0

These results correspond to the results of the classification carried out with the eight dimensional feature vectors based on subband level 4, however now the dimension of the feature vectors is 17 instead of 8.

5. Classification of Noisy Patterns

Reliability of the classifier may be indicated by the geometric margin, but the robustness should be studied via perturbation of the time signals. Noise of normal distribution with mean of unity and standard deviation of σ was used to multiply the original time signals, randomly, and classification of these noisy signals has been done using the same SVM classifiers trained with noiseless signals.

```
<<Statistics‘ContinuousDistributions‘
    ListPlot[Map[(#
    Random[NormalDistribution[1.,0.1]])&,Table[1,{100}]],
Frame → True,PlotJoined → True,Axes → None,AspectRatio → 0.4];
```
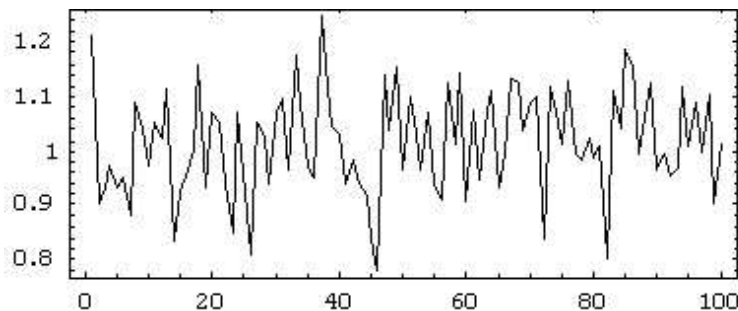
Fig. 5. Random values of the multiplier in case of $\sigma = 0.1$

```
ListPlot[Map[(# Random[NormalDistribution[1.,0.1]])&,Drop[ts,70000]],
Frame → True,PlotJoined → True,Axes → None,FrameLabel →
{"Samples","Blood Flow AU"},AspectRatio → 0.4];
```
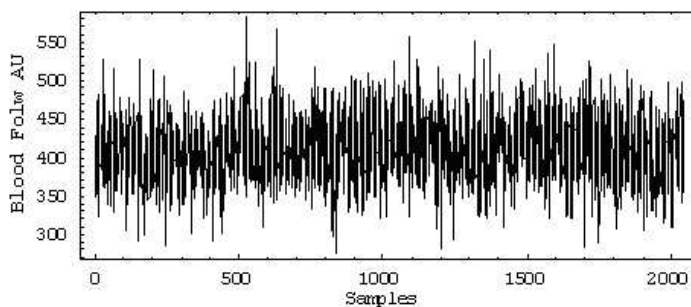

Fig. 6. Noisy time signal, $\sigma = 0.1$

Classifications based on the subband method are more robust, less sensitive for the perturbation of the time signals, and result less misclassified patterns than that based on the traditional feature extraction.

6. Conclusions

In this article an improved feature extraction method is presented for classification of time series. The main advantage of this feature extraction method is that it can be efficiently used for classification even if the number of measurements is limited. This method results in low dimensional feature vectors. Combining the method with support vector classification leads to a very robust and reliable classification procedure. These advantages of the introduced feature extraction were successfully

Table 3. The number of misclassified patterns, using different σs in case of the traditional and two different subband feature vectors.

Standard deviation of noise	Traditional feature extraction	Subband method with level 3	Subband method with level 5
0.01	0	0	0
0.02	7	0	0
0.03	9	0	0
0.04	11	0	0
0.05	11	0	0
0.10	-	0	0
0.20	-	2	0
0.30	-	5	0
0.40	-	8	4

illustrated by solving the classification problem of CBF signals, which could be solved only partially by other traditional methods [12].

Acknowledgements

References

[1] LACZA, Z.–ERDŐS, B.–GÖRLACH, C.–HORTOBÁGYI, T.–SÁNDOR, P.–WAHL, M.–BENYÓ, Z., The cerebrocortical microcirculatory effect of nitric oxide synthase blockade is dependent upon baseline red blood cell flow in the rat. *Neuroscience Letters*, **291** (2000), pp. 65-68.

[2] LENZSÉR, G.–HERMÁN, P.–KOMJÁTI, K.–SÁNDOR, P.–BENYÓ, Z., Nitric oxide synthase blockade sensitizes the cerebrocortical circulation to thromboxane-induced CBF oscillations. *Journal of Cerebral Blood Flow and Metabolism*, **23** (2003), pp. 88.

[3] BENYÓ, B.–LENZSÉR, G.–PALÁNCZ, B., Characterization of the Temporal Pattern of Cerebral Blood Flow Oscillations, Proc. of 2004 International Joint Conference on Neural Networks, Budapest, Hungary, July 25-29, 2004, pp. 2467-2470.

[4] BENYÓ, B.–SOMOGYI, P.– PALÁNCZ, B., Characterization of Cerebral Blood Flow Oscillations Using Different Classification Methods, IFAC'05 Conference, Prague, July 4-8, 2005.

[5] GOSWAMI, J. C.–CHAN, A. K., *Fundamentals of Wavelets. Theory, Algorithms, and Application*. John Wiley, New York, 1999.

[6] CHOE, H.C, Ph.D. dissertation, Texas A&M University, College Station, Texas, USA, 1997.

[7] PALÁNCZ, B., Wavelets and their Application to Digital Image Processing, *Publications in Geomatics*, **7** (2004), pp.59-68.

[8] ABOUFADEL–SCHLICKER, *Discovering Wavelets*, John Wiley, New York, 1999.

[9] ZHANG, L.–ZHOU, W.–JIAO, L., Hidden Space Support Vector Machines, *IEEE Trans. on Neural Networks*, **15/6** (2004), pp.1424-1434.

[10] CRISTIANINI, N.–SHAWE-TAYLOR, J., *An Introduction to Support Vector Machines and other kernel learning methods*, Cambridge University Press, 2003.

[11] PALÁNCZ, B., Support Vector Classifier via Mathematica, *Wolfram Research Mathematica Information Center*, e-publication, http//library.wolfram.com/infocenter/mathsource/5293 2004.

[12] DUDA, R.O.–HART, P. E.–STORK, D.G., *Pattern Classification*, John Wiley, New York, second ed., 2001.