TWO ALGORITHMS FOR DESIGNING OPTIMAL REDUCED-BIAS DATA-DRIVEN TIME-FREQUENCY DETECTORS

C. Richard, R. Lengellé

Laboratoire LM2S - Université de Technologie de Troyes
12, rue Marie Curie - BP 2060
10010, Troyes cedex, FRANCE

ABSTRACT

Designing time-frequency detectors from training data is potentially of great benefit when few a priori information on the non stationary signal to be detected is available. However, achieving good performance with data-driven detectors requires matching their complexity to the available amount of training samples: receivers with a too large number of adjustable parameters often exhibit poor generalization performance whereas those with an insufficient complexity cannot learn all the information available in the set of training data. In this paper, we present two methods which provide powerful tools for tuning the complexity of time-frequency detectors and improving their performance. These procedures may offer an helpful support for designing efficient detectors from small training sets, in applications of current interest such as biomedical engineering and complex systems monitoring.

1. INTRODUCTION

Cohen's class time-frequency (TF) representations are potentially useful for detection in applications of current interest such as microemboli diagnosis [1], or sleep EEG analysis [2], due to the need for dealing with non-stationary signals. Most of the TF based receivers which were proposed are linear structures operating in the TF domain, and are merely equivalent to quadratic detectors usually defined in the time domain. In [3], Sayeed and Jones also proposed a promising TF-based quadratic theory: they identify several scenarios in which detectors are optimum and exploit the structure of the representations. Recently, these authors have extended the scope of their theory to generalized joint signal representations [4].

All these approaches require prior knowledge of the event to be detected whereas phenomena are complex and poorly understood in many applications. In this context, several authors proposed to design TF detectors directly from labeled training data [2, 5-8]. Nevertheless, it is well known in Pattern Recognition that data-driven receivers often have a large bias, particularly when the number of training samples is small against the dimension of data. This experimental evidence has been theoretically studied by Vapnik and Chervonenkis, who exhibited links between the generalization performances of receivers, their complexity, and the size of the training set [9].

In this paper, we present two methods based on the Structural Risk Minimization (SRM) principle proposed by Vapnik [10]. In both cases, the strategy consists in a reduction of the effective dimension of the detector, which can yield a substantial improvement in its performance. The paper is organized as follows. First, we briefly describe an efficient method of designing TF-based detectors from training data. Next, we discuss the issue of obtaining reduced-bias TF-based receivers. Then, two methods which are reminiscent of procedures applied after neural networks training are considered. Finally, an example illustrates the efficiency of these approaches.

2. DATA-DRIVEN TIME-FREQUENCY DETECTORS

It is shown in [3] that linear TF-based detectors are optimal for a variety of composite hypothesis testing scenarios. However, we will focus on the hypothesis testing problem (1) because the scope of our approach can easily be extended, using the procedure proposed in [7]:

$$\begin{cases} \text{if } \lambda(\underline{X}; v) = \underline{V}^T \underline{X} \geq \eta \quad \text{then } H_1 \\ \text{else } H_0, \end{cases} \tag{1}$$

where

$$\underline{X} = \left[W_x(t_1,f_1)\ W_x(t_2,f_1) \ldots W_x(t_d,f_d)\right]^T$$
$$\underline{V} = \left[v(t_1,f_1)\ v(t_2,f_1) \ldots v(t_d,f_d)\right]^T. \tag{2}$$

W_x denotes the Wigner-Ville distribution of the discrete-time observation x over the interval $[t_1, t_d]$. The bi-dimensional function v is a TF reference to be determined using the *a priori* knowledge of some observations, conditionally to H_0 and H_1.

The design of a detector from training data consists in finding the optimum ($\underline{V}$; η) in the sense of a pre-selected criterion and for a given data set. In [2, 5], it is shown that the maximization of any scatter criterion f, depending only on the 1^{st} and 2^{nd} order moments of λ (Fisher, SNR, ...) conditionally to H_0 and H_1, leads us to the solution $\underline{V}_\alpha$ which satisfies:

$$[\alpha\Sigma_0 + (1-\alpha)\Sigma_1]\, \underline{V}_\alpha = (\underline{M}_1 - \underline{M}_0), \qquad (3)$$

where $\underline{M}_i$ and Σ_i respectively denote the 1^{st} and 2^{nd} order moments of $\underline{X}$, conditionally to H_i. It can be shown that the parameter α is a member of [0, 1] and only depends on the criterion f.

In Eq. (3), the effect of the criterion f appears only in α. As a consequence, one can optimize this parameter in order to minimize an estimation of the probability of error E. The resulting receiver obviously offers a better performance than detectors determined via the maximization of the SNR and of the Fisher criterion, which correspond to $\alpha = 1$ and $\alpha = P\{H_0\}$, respectively [2, 11]. As a conclusion, this method allows to determine the optimum receiver in the sense of the best criterion f depending only on the 1^{st} and 2^{nd} order moments of the statistic λ. This criterion is never set up.

3. OPTIMIZATION OF TIME-FREQUENCY BASED DETECTION STRUCTURES

3.1. Complexity regularization

Achieving good performances with detectors designed from training samples requires matching their complexity to the amount of available data: receivers with a too large number of adjustable parameters may exhibit poor generalization performances, whereas those with an insufficient complexity may not be able to learn the training examples. In between, there is an optimal complexity which yields the lowest probability of error E for a given size of the training set. In this paper, the approach that we present consists in jointly optimizing α and reaching this compromise with the complexity of the detector, in order to minimize an estimation of E. Given α, we propose to control the complexity of $\underline{V}_\alpha$ as follows.

3.2. Principle of Optimal Brain Damage (OBD)

One common way of reducing the complexity of the receiver $\underline{V}_\alpha$ is to set some of its components to zero and thereby reduce the number of free parameters. From Eq. (3), we define the best candidates for pruning as those which minimize the increase of the Mean Square Error (MSE) defined as follows:

$$MSE_\alpha = \|\Sigma_\alpha . \underline{V}_\alpha - \underline{M}\|^2 \qquad (4)$$

where

$\Sigma_\alpha = \alpha\, \Sigma_0 + (1-\alpha)\, \Sigma_1$ and $\underline{M} = \underline{M}_1 - \underline{M}_0$.

Since the decrease in complexity should be achieved at the smallest possible expense in MSE increase, it can be shown that the pruning process must be performed in a basis of normalized eigenvectors Φ of Σ_α [11]. Note that Σ_α can be diagonalized because this matrix is symmetric. In this basis, the increase $\Delta MSE_{i,\alpha}$ due to setting the i^{th} component of $\underline{V}_\alpha$ to zero is of the form:

$$\Delta MSE_{i,\alpha} = [\mu_i\, \underline{V}_{i,\alpha}]^2 \qquad (5)$$

where

$\underline{V}_{i,\alpha} = (\underline{\Phi}_i)^T\, \underline{V}_\alpha$.

In Eq. (5), μ_i (res. Φ_i) denotes the i^{th} eigenvalue (res. eigenvector) of Σ_α, and $\underline{V}_\alpha$ satisfies Eq. (3).

Consequently, the components of $\underline{V}_\alpha$ which correspond to the smallest increases ΔMSE_α are good candidates for elimination. This approach is reminiscent of *Optimal Brain Damage*, a weight pruning procedure commonly applied after neural networks training [12].

3.3. Principle of Weight Decay (WD)

The complexity of the receiver can also be controlled through an additional penalty term to be simultaneously minimized with MSE_α. In [10], Vapnik suggest the minimization of the new cost function $MSE_\alpha + \gamma\, \|\underline{V}_\alpha\|^2$. In the case of linear detectors, this operation is equivalent to pull the

Fig.1: WV of the signal s

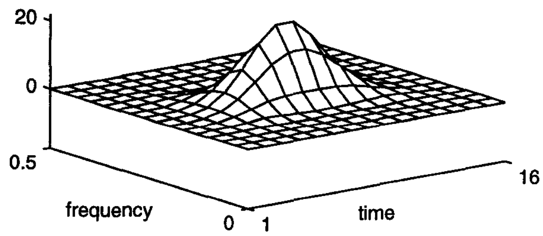


Fig.2: Biased detector (16^2 parameters)

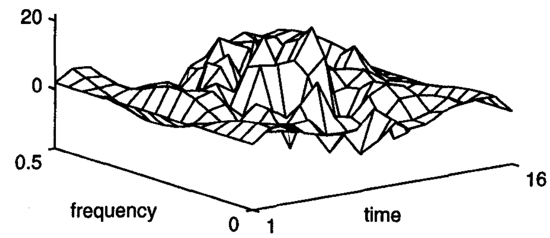


Fig.3: Reduced-bias detector (OBD, 7 parameters)

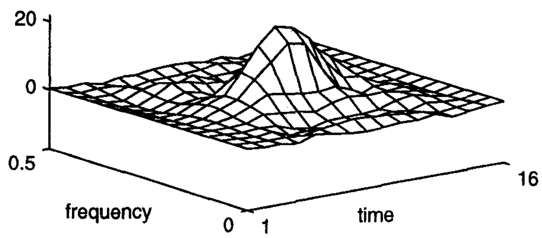
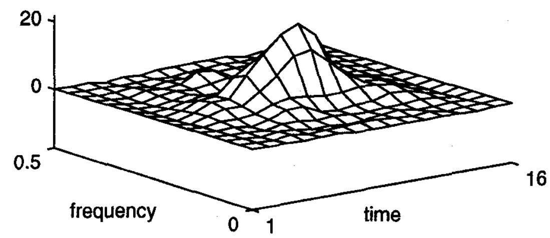


Fig.4: Reduced-bias detector (WD, 12 significant parameters)



components $\underline{V}_{i,\alpha}$ to zero predominantly along the principal directions of Σ_α associated with small eigenvalues since we have, in the basis of eigenvectors Φ:

$$\underline{V}_{i,\alpha} = \mu_i^2 / \left[\mu_i^2 + \gamma\right] \underline{V}_{i,\alpha}, \qquad (6)$$

where μ_i, Φ_i and $\underline{V}_\alpha$ are defined as in Eq. (5).

As a conclusion, the effect of the penalty term $\|\underline{V}_\alpha\|^2$ can be compared to that of a pruning procedure, such as the one introduced in Section 3.2.

4. EXPERIMENTAL RESULTS

Several experiments of blind detector design from training data were conducted in order to illustrate the efficiency of our approach. In the case of detecting the presence or absence of
$s(k) = \sin(0.44\pi k + \theta)\times(1 - \cos(2\pi k/15))$, $k \in [0, 15]$, in zero mean white Gaussian noise (RSB = -6 dB), with phase θ a uniform random variable, the optimal receiver is known to be to the inner product of the Wigner-Ville distribution of the signal s (Fig.1) with that of the observation x. The design of TF-based detectors was conducted with 260 realizations of the hypotheses H_0 and H_1, in such way that the training set was sparse compared with the problem dimension (16^2+1). The TF reference resulting from the direct resolution of Eq. (3) is shown in Fig.2: the presence of the signal component is not very apparent because few training examples were available. The methods proposed in Sections 3.2 and 3.3 were used to design reduced-bias TF-based detectors. The obtained references closely resembles the Wigner-Ville representation of s, as shown in Figs. 3 and 4.

The generalization performance of these detectors were estimated by applying them to 2000 realizations each of signal present and signal absent. Using the quadrature matched filter, which is the optimal detector, the generalization error was 20.90%. The performance of the receiver resulting from the resolution of Eq.(3) was 24.10%. This result must be compared to 21.87% (res. 20.95%) obtained with the reduced-bias detector, when using the *Optimal Brain Damage* (res. *Weight Decay*) procedure.

As a conclusion, these experiments clearly demonstrate the ability of the proposed methods to closely approach the performances of the optimal quadratic detector, even if the size of the training set is relatively small compared to the problem dimension.

5. CONCLUSION

In this paper, we have presented a method of designing time-frequency detectors which requires no prior knowledge of the event to be detected. The receivers are directly derived from training data and theoretically perform better than those obtained via the maximization of the Fisher criterion or the signal to noise ratio. However, it is well known in Pattern Recognition that the performance of classifiers strongly depends on their complexity and on the number of available training data. Then, we have also introduced two procedures which provide powerful tools for tuning the complexity of data-driven time-frequency detectors and improving their performance. Moreover, they are computationally more efficient than the approach proposed in [13]. Finally, we have successfully experimented these methods on simulated data.

6. REFERENCES

[1] B. S. KRONGOLD, A. M. SAYEED, M. MOEHRING, J. A. RITCEY, M. SPENCER, and D. L. JONES, "Quadratic Time-Scale Detection of Microemboli in Flowing Blood with Doppler Ultrasound," *SPIE's 42nd Meeting*, 1997.

[2] C. RICHARD and R. LENGELLE, "Joint Time and Time-Frequency Optimal Detection of K-complexes in Sleep EEG," *Computers and Biomedical Research*, vol. 31, pp. 209-229, 1998.

[3] A. M. SAYEED and D. L. JONES, "Optimal Detection using Bilinear Time-Frequency and Time-Scale Representations," *IEEE Transactions on Signal Processing*, vol. 43, pp. 2872-2883, 1995.

[4] A. M. SAYEED and D. L. JONES, "Optimal Quadratic Detection and Estimation using Generalized Joint Signal Representation," *IEEE Transactions on Signal Processing*, vol. 44, pp. 2959-2970, 1996.

[5] C. RICHARD and R. LENGELLE, "Une Nouvelle Approche pour la Détection Linéaire Optimale dans le Plan Temps-Fréquence," *Seizième Colloque GRETSI*, pp. 659-662, 1997.

[6] B. SAMIMY, G. RIZZONI, A. M. SAYEED, and D. L. JONES, "Design of Training Data-Based Quadratic Detectors with Application to Mechanical Systems," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1767-1770, 1996.

[7] A. M. SAYEED, "Data-Driven Time-Frequency and Time-Scale Detectors," *SPIE's 42nd Meeting*, 1997.

[8] D. L. JONES and A. M. SAYEED, "Blind Quadratic and Time-Frequency based Detectors from Training Data," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1033-1036, 1995.

[9] V. N. VAPNIK and A. Y. CHERVONENKIS, "On the Uniform Convergence of Relative Frequencies of Events to their Probabilities," *Theory of Probabilities and its Applications*, vol. 16, pp. 264-280, 1971.

[10] V. N. VAPNIK, Estimation of Dependences Based on Empirical Data. New York, NY: Springer Verlag, 1982.

[11] C. RICHARD and R. LENGELLE, "Data-Driven Design and Complexity Control of Time-Frequency Detectors," *submitted to Signal Processing*, 1998.

[12] Y. LE CUN, J. DENKER, and S. SOLLA, "Optimal Brain Damage," in Advances in Neural Information Processing 2, D. S. TOURETZKY, Ed. San Mateo: Morgan Kofmann, 1989, pp. 598-605.

[13] C. RICHARD and R. LENGELLE, "Structural Risk Minimization for Reduced-Bias Time-Frequency-Based Detectors Design," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2397-2400, 1998.