SUPPORT VECTOR-BASED ONLINE DETECTION OF ABRUPT CHANGES

Frédéric Desobry and Manuel Davy

IRCCyN, UMR CNRS 6597,
1 rue de la Noë - BP92101
44321 Nantes Cedex 3 - France
{Frederic.Desobry,Manuel.Davy}@irccyn.ec-nantes.fr

ABSTRACT

We present a machine learning technique aimed at detecting abrupt changes in a sequence of vectors. Our algorithm requires a Mercer kernel together with the corresponding feature space. A stationarity index is designed in the feature space, and consists of comparing two circles corresponding to two ν-SV novelty detectors via a Fisher-like ratio. An abrupt change corresponds to a large distance between the circles centers (w.r.t. their radii). We show that the index can be computed in the input space, and simulation results show its efficiency in front of real data.

1. INTRODUCTION

Abrupt changes detection in signals is a much studied problem, and various approaches have been proposed. Some rely on the knowledge of a signal statistical model; Generalized Likelihood Ratio (GLR) techniques [1] and Bayes detection theory have excellent performance. However, in some applications, it may be difficult to design an accurate and tractable statistical model, and model-free approaches need then be considered. In this paper, we propose a model-free, machine learning based online algorithm for abrupt changes detection in signals.

Our algorithm is two-step. First, informative descriptors (or vectors) localised in time, denoted $\vec{x}_t$, are extracted online from the signal. These can be cepstral coefficients computed on a sliding window, short-time Fourier transforms (STFTs), etc. Second, we define a kernel-based online stationarity index $I(t)$ computed in the descriptors space (or input space, denoted $\mathcal{X}$), and geometrically defined in a feature space $\mathcal{F}$ induced by a Mercer kernel $k(\cdot,\cdot)$[1]. Roughly, $I(t)$ is computed as follows. A kernel $k(\cdot,\cdot)$ is selected. At time t, a first ν-Support Vector (SV) novelty detector is trained over the m_1 last descriptors $\boldsymbol{x}_1 = (\vec{x}_{t-m_1}, \cdots, \vec{x}_{t-1})$, yielding a decision region $\mathcal{R}_1$ in $\mathcal{X}$. $\mathcal{R}_1$ is such that a vector $\vec{x}$ is considered similar to $\boldsymbol{x}_1$ iff $\vec{x} \in \mathcal{R}_1$. Next, a second ν-SV novelty detector is trained over the m_2 future descriptors $\boldsymbol{x}_2 = (\vec{x}_t, \cdots, \vec{x}_{t+m_2-1})$, yielding $\mathcal{R}_2$. The regions $\mathcal{R}_1$ and $\mathcal{R}_2$ are representative of the probability density functions (pdfs) which generated the sets $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Thus, comparing the geometries and locations of $\mathcal{R}_1$ and $\mathcal{R}_2$ is a robust way to compare $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. We show that this can be easily done in the feature space, and that it can be computed in the input space using $k(\cdot,\cdot)$.

The remainder of this paper is organized as follows. In Section 2, we recall basic elements about ν-SV novelty detection. In Section 3, our algorithm is described. In particular, we explain how the stationarity index is defined in the feature space, and computed in the input space. Simulation results are presented in Section 4, and some conclusions and perspectives are given in Section 5.

[1]The relation between Mercer kernels and feature spaces is exposed in, e.g., [2].

2. ν-SV NOVELTY DETECTION

We assume a set of m training points $\boldsymbol{x} = (\vec{x}_1, \cdots, \vec{x}_m)$ is available in the input space $\mathcal{X}$. We define a learning algorithm

$$\begin{array}{rccc} \mathcal{A} & : & \cup_{m=1}^{\infty}\mathcal{X}^m & \longrightarrow & \mathcal{H} \\ & & \{\vec{x}_1,\ldots,\vec{x}_m\} & \longmapsto & \mathcal{A}(\boldsymbol{x}) \end{array}$$

where $\mathcal{H}$ is a *hypothesis space* of indicator functions $\mathbf{I}_{\mathcal{R}}$ such that $\mathbf{I}_{\mathcal{R}}(\vec{x}) = 1$ if $\vec{x} \in \mathcal{R}$ and 0 otherwise for any subsets $\mathcal{R}$ in $\mathcal{X}$. Next, we define a mapping $\boldsymbol{\phi}$ from $\mathcal{X}$ to a so-called *feature space* $\mathcal{F}$:

$$\begin{array}{rccc} \boldsymbol{\phi} & : & \mathcal{X} & \longrightarrow & \mathcal{F} \\ & & \vec{x} & \longmapsto & \mathbf{x} = \boldsymbol{\phi}(\vec{x}) \end{array}$$

We assume $\mathcal{F}$ is endowed with a dot product $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$, and restrict $\mathcal{H}$ to indicator functions on decision regions of the form $\mathcal{R} = \{\mathbf{x} : f_{\boldsymbol{x}}(\vec{x}) > 0\}$ where the decision function is defined as

$$\begin{array}{rccc} f_{\boldsymbol{x}} & : & \mathcal{X} & \longrightarrow & \mathbb{R} \\ & & \vec{x} & \longmapsto & f_{\boldsymbol{x}}(\vec{x}) = \langle \mathbf{x}, \mathbf{w} \rangle - \rho \end{array} \tag{1}$$

In eq. (1), $\mathbf{w}$ is a linear combination of mapped training points,

$$\mathbf{w} = \sum_{i=1}^{m} \alpha_i \mathbf{x}_i, \quad \text{with}\, \alpha_i \geq 0 \quad \forall i = 1, \ldots, m. \tag{2}$$

The parameters $\mathbf{w}$ and ρ, that completely define $\mathcal{A}(\boldsymbol{x})$ for a given training set $\boldsymbol{x}$, are determined by solving

$$\begin{array}{l} \max\limits_{\mathbf{w},\boldsymbol{\xi},\rho} -\frac{1}{2}\|\mathbf{w}\|^2 - \frac{1}{\nu m}\sum\limits_{i=1}^{m}\xi_i + \rho \\ \text{subject to } \langle \mathbf{w}, \mathbf{x}_i \rangle \geq \rho - \xi_i, \quad \xi_i \geq 0. \end{array} \tag{3}$$

where $0 \leq \nu \leq 1$. In the dual formulation, we obtain directly α_i, $i = 1, \ldots, m$ by minimizing (w.r.t. α_i's and ρ)

$$W = +\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

$$\text{subject to } 0 \leq \alpha_i \leq \frac{1}{\nu m} \text{ for } i = 1, \ldots, m \quad \text{and} \sum_{i=1}^{m}\alpha_i = 1.$$

All training points $\vec{x}_i$ for which $f_{\boldsymbol{w}}(\vec{x}_i) \leq 0$ are called *support vectors* (SVs); these are the only points for which SV algorithms yield $\alpha_i \neq 0$ in eq. (2), thus the SVs alone determine $f_{\boldsymbol{w}}(\cdot)$. SVs are divided into two sets: *margin SVs*, for which $f_{\boldsymbol{w}}(\vec{x}_i) = 0$, and *non-margin SVs*, for which $f_{\boldsymbol{w}}(\vec{x}_i) < 0$. It is proved in [3] that ν is an upper bound on the fraction of non-margin SVs, and a lower bound on the fraction of support vectors; in addition, ν is asymptotically equal to both the fraction of SVs and the fraction of non-margin SVs with probability 1, under mild conditions on the probability distribution generating the data.

Finally, we note that we need never compute the mapping $\boldsymbol{\phi}(\vec{x})$: rather, by eq. (2), it follows that $\langle \mathbf{w}, \mathbf{x}_j \rangle$ can be computed using only the dot product function, $k(\vec{x}_i, \vec{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$. A kernel k represents a dot product in some feature space if it fulfills the Mercer conditions [4]. These conditions are satisfied for a wide range of kernels, including Gaussian radial basis functions,

$$k\left(\vec{x}_i, \vec{x}_j\right) = \exp\left(-\frac{1}{2\sigma^2}\left\|\vec{x}_i - \vec{x}_j\right\|^2\right). \tag{4}$$

3. ONLINE ABRUPT CHANGES DETECTION ALGORITHM

In this section, we introduce our main result, namely an algorithm aimed at detecting online abrupt changes in the distribution of vectors $\vec{x}_t$ observed sequentially in time. The sequence $\vec{x}_t$ can be produced directly by a system. In signal processing applications, it typically results from a preprocessing step aimed at extracting descriptors from a signal at each time instant t.

3.1. Algorithm description

Consider at time t two subsets $\boldsymbol{x}_1 = (\vec{x}_{t-m_1}, \cdots, \vec{x}_{t-1})$ and $\boldsymbol{x}_2 = (\vec{x}_t, \cdots, \vec{x}_{t+m_2-1})$ of size m_1 and m_2. Each of these subsets are used to train independently ν-SV novelty detectors, yielding parameters $\mathbf{w}_1, \rho_1$ and $\mathbf{w}_2, \rho_2$, and decision regions $\mathcal{R}_1$ and $\mathcal{R}_2$. The idea underlying our abrupt changes detector is that a sudden change at time t in the distribution of vectors $\vec{x}_t$ may result in different locations/geometries of $\mathcal{R}_1$ and $\mathcal{R}_2$ (as depicted in the input space in fig. 2): in practice, at each time t, we want to build an index $I(t)$ that reflects the dissimilarity between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ via a measure of the dissimilarity between $\mathcal{R}_1$ and $\mathcal{R}_2$. (The computation of $I(t)$ for a given t is described in the next subsection). Once $I(t)$ is computed, the subsets $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are updated at time $t+1$. The corresponding parameters $\mathbf{w}_1, \rho_1$ and $\mathbf{w}_2, \rho_2$ are also updated using the online ν-SV novelty detection technique presented in [5]. This online technique avoids computing the ν-SV novelty detection parameters $\mathbf{w}, \rho$ from scratch at each time t. Abrupt changes are finally detected whenever the index $I(t)$ peaks or is over a threshold η (this is similar to most abrupt changes detection techniques). In the next subsection, we propose an original stationarity index induced by geometrical considerations in the feature space.

3.2. Geometry in the feature space

Consider a normalized Mercer kernel such that $k(\vec{x}, \vec{x}) = 1,\ \forall \vec{x}$. The mapped training vectors $\mathbf{x}_t = \phi(\vec{x}_t)$ are located on the hypersphere $\mathcal{S}$ centered at the origin of $\mathcal{F}$ (denoted $\mathbf{0}$), with radius 1. Given a training set, say $\boldsymbol{x}_1$, the optimisation problem given eq. (3) admits the following geometrical interpretation. The parameters $\mathbf{w}_1, \rho_1$ define a hyperplane $\mathcal{W}_1$ in $\mathcal{F}$, orthogonal to $\mathbf{w}_1$ and distant of $\rho_1/||\mathbf{w}_1||$ from $\mathbf{0}$ (see fig. 1). $\mathcal{W}_1$ separates the mapped training vectors from $\mathbf{0}$. In $\mathcal{F}$, the mapped region $\mathcal{R}_1$ is the portion of $\mathcal{S}$ that is delimited by $\mathcal{W}_1$, and that is located opposite $\mathbf{0}$. Its boundary is a hypercircle $\mathcal{C}_1$ in $\mathcal{F}$. The hypersphere $\mathcal{S}$ radius being one, $\mathcal{C}_1$ is entirely defined by $\mathbf{w}_1, \rho_1$. Similarly, for $\boldsymbol{x}_2$, solving eq. (3) yields an hyperplane $\mathcal{W}_2$ and a hypercircle $\mathcal{C}_2$ defined by $\mathbf{w}_2, \rho_2$.

In the feature space, the shapes of the mapped regions $\mathcal{R}_1$ and $\mathcal{R}_2$ are simple (their boundaries are hypercircles $\mathcal{C}_1$ and $\mathcal{C}_2$). They can be fully compared from their center locations and radii, which can be computed from $\mathbf{w}_1, \rho_1$ and $\mathbf{w}_2, \rho_2$.

A simple way to compare $\mathcal{C}_1$ and $\mathcal{C}_2$ (i.e., $\mathcal{R}_1$ and $\mathcal{R}_2$), and to build $I(t)$, consists of considering an interclass/intraclass ratio in $\mathcal{F}$, such as

$$I(t) = \frac{\text{distance between circle centers}}{\text{radius of } \mathcal{C}_1 + \text{radius of } \mathcal{C}_2}$$

If $\mathbf{w}_1 = \mathbf{w}_2$, which means that both training sets are located around the same position, then $I(t) = 0$. This is what we expect in such situations. Let now assume $\mathbf{w}_1 \neq \mathbf{w}_2$. Then, there exists at least at least one triplet $(\mathbf{p}_1, \mathbf{p}_2, \mathcal{C})$ such that:

$$\begin{cases} \mathbf{p}_1 \in \mathcal{S} \cap \mathcal{W}_1 \\ \mathbf{p}_2 \in \mathcal{S} \cap \mathcal{W}_2 \\ \mathcal{C} \text{ is a circle of radius 1 with center } \mathbf{0} \\ \mathbf{c}_1, \mathbf{c}_2, \mathbf{p}_1, \text{ and } \mathbf{p}_2 \text{ are all located on } \mathcal{C}. \end{cases} \tag{5}$$

where $\mathbf{c}_1$ (respectively $\mathbf{c}_2$) are the points located at the intersection between the line passing through $\mathbf{0}$ and oriented along $\mathbf{w}_1$ (respectively $\mathbf{w}_2$) and $\mathcal{S}$, see fig. 1.

$\mathbf{c}_1$ and $\mathbf{c}_2$ are the (geometrical) centers of the region occupied by the feature data, and $\mathbf{p}_1$ and $\mathbf{p}_2$ possess properties similar to mapped margin support vectors as they lie on the separating hyperplane. Yet all the 4 points possibly do not have pre-images in the input space. However, $I(t)$ can be expressed in terms of dot products in the feature space, i.e., as kernels in the input space. We note that $I(t)$ can be written

$$I(t) = \frac{|\mathbf{c}_1\widehat{\mathbf{c}_2}|}{|\mathbf{c}_1\widehat{\mathbf{p}_1}| + |\mathbf{c}_2\widehat{\mathbf{p}_2}|} \tag{6}$$

where $\widehat{\ }$ denotes the circle arc. As $\mathcal{S}$ is of radius 1, the circle arc is also the circle angle between $\mathbf{c}_1$ and $\mathbf{c}_2$, and

$$\mathbf{c}_1\widehat{\mathbf{c}_2} = \widehat{\mathbf{c}_1\mathbf{0}\mathbf{c}_2} = \cos^{-1}\left(\frac{< \mathbf{w}_1, \mathbf{w}_2 >}{||\mathbf{w}_1||.||\mathbf{w}_2||}\right) \tag{7}$$

Using eq. (2), we have

$$\frac{< \mathbf{w}_1, \mathbf{w}_2 >}{||\mathbf{w}_1||.||\mathbf{w}_2||} = \frac{\sum_{i,j} \alpha_{1i}\alpha_{2j}k(\vec{x}_{1i}, \vec{x}_{2j})}{(\sum_{i,j} \alpha_{1i}\alpha_{1j}k(\vec{x}_{1i}, \vec{x}_{1j}))^{0.5}(\sum_{i,j} \alpha_{2i}\alpha_{2j}k(\vec{x}_{2i}, \vec{x}_{2j}))^{0.5}}$$

where α_{1i} (resp. α_{2i}) correspond to $\mathbf{w}_1$ (resp. $\mathbf{w}_2$) and $\vec{x}_{1i}$ (resp. $\vec{x}_{2i}$). Using similar computations, we can express the circle arc between $\mathbf{c}_1$ and $\mathbf{p}_1$ as

$$\mathbf{c}_1\widehat{\mathbf{p}_1} = \cos^{-1}\left(\frac{\rho_1}{(\sum_{i,j} \alpha_{1i}\alpha_{1j}k(\vec{x}_{1i}, \vec{x}_{1j}))^{0.5}}\right)$$

We obtain a similar result for $\mathbf{c}_2\widehat{\mathbf{p}_2}$. This shows that $I(t)$ is a function that can be computed in $\mathcal{X}$ using $k(\cdot, \cdot)$ only.

Remark– If we use a positive kernel $k(\vec{x}_i, \vec{x}_j) > 0\ \forall (i, j)$, all mapped data lie in the same orthant, and the angle $\widehat{\mathbf{c}_i\mathbf{0}\mathbf{p}_i}$ $(i = 1, 2)$

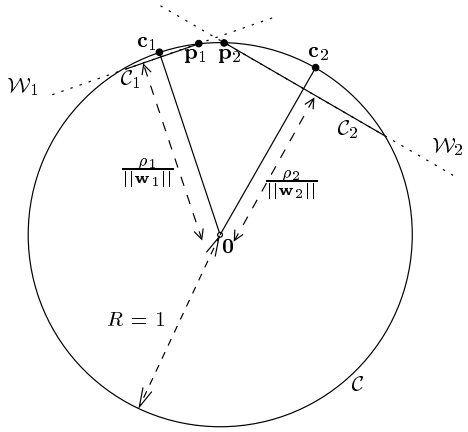

Fig. 1. In the feature space $\mathcal{F}$, the training data are mapped on a hypersphere $\mathcal{S}$ with radius 1 and center $\mathbf{0}$. The ν-SV novelty detector related to $\boldsymbol{x}_1$ (resp. $\boldsymbol{x}_2$) yields a hypercircle $\mathcal{C}_1$ (resp. $\mathcal{C}_2$) in the hyperplane $\mathcal{W}_1$ (resp. $\mathcal{W}_2$).

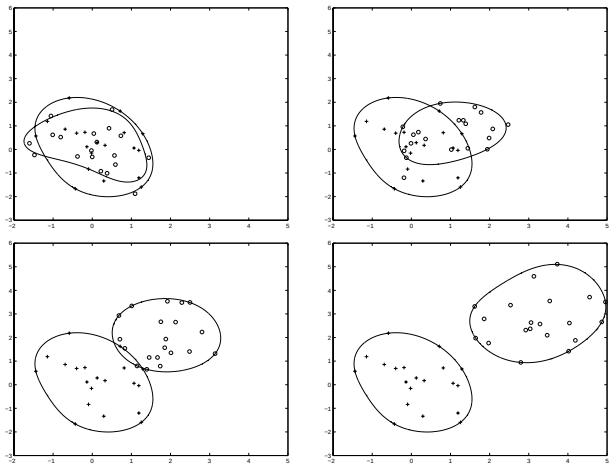

Fig. 2. In the input space, evolution of the separating curves yielded by the two ν-SV novelty detectors when the mean of one of the two subsets change (top left, top right, bottom left, bottom right). Two surfaces with a small overlap correspond to a change.

cannot exceed $\pi/2$. The angle $\widehat{\mathbf{c}_1\mathbf{0}\mathbf{c}_2}$ is likely to be lower than $\pi/2$, especially if there is no abrupt change. In this case, the cosine being bijective on $[0;\pi/2[$, we can consider the modified index

$$\tilde{I}(t) = \frac{\cos(\widehat{\mathbf{c}_1\mathbf{c}_2})}{\cos(\widehat{\mathbf{c}_1\mathbf{p}_1}) + \cos(\widehat{\mathbf{c}_2\mathbf{p}_2})} = \frac{< \mathbf{w}_1/\rho_1, \mathbf{w}_2/\rho_2 >}{||\mathbf{w}_1||/\rho_1 + ||\mathbf{w}_2||/\rho_2} \quad (8)$$

3.3. Discussion

From eq. 6, we see that $I(t)$ is indeed a Fisher-like ratio. Hence, abrupt changes will be detected whenever they correspond to an abrupt change of means (due to the term $\mathbf{c}_1\widehat{\mathbf{c}_2}$) w.r.t. to the scale of the data (given by $\mathbf{c}_1\widehat{\mathbf{p}_1}$ and $\mathbf{c}_2\widehat{\mathbf{p}_2}$). Note that $I(t)$ is robust to outliers $\vec{x}_t$ which are not due to abrupt changes, because ν-SV novelty detection rejects such outliers. Though based on the same fundamentals as the Rayleigh coefficient used in Kernel Fisher Discriminant (KFD) analysis [2], our coefficient $I(t)$ is different: in KFD, the Rayleigh coefficient is maximized w.r.t. the projection direction $\mathbf{w}$ whereas in our technique $\mathbf{w}$ is yielded by the ν-SV novelty detectors.

Another approach based on SV classification could be considered, where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are seen as training sets of two different classes. One could train a SVM classifier using these two sets, and consider the margin as the stationarity index. However, this approach suffers from several drawbacks. First, when no abrupt change occurs, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are mixed; thus designing a classifier is somewhat artificial, and meaningless. Second, vectors considered as outliers in ν-SV classification are different from those considered as outliers in the two ν-SV novelty detectors. The way the former are selected in one set (e.g., $\boldsymbol{x}_1$) is governed by the other set (e.g., $\boldsymbol{x}_2$), and not by their intrisic behavior w.r.t. the process $\vec{x}_t$. Third, the computational burden is higher with SV classification because training with $m_1 + m_2$ vectors (in $\boldsymbol{x}_1 \cup \boldsymbol{x}_2$) costs more than training independently with m_1 and m_2 vectors.

Finally, the approach in [6] addresses a related but different problem, namely early abnormality detection (where no detection delay is tolerated). In [6], a candidate $\vec{x}_t$ is tested with respect to the novelty detector trained on the set $\{\vec{x}_{t-m}, ..., \vec{x}_{t-1}\}$. A stationarity index is built with the output of the detector:

$$I_o(t) =< \mathbf{w}_o(t), \phi(\vec{x}_t) > -\rho_o(t) \quad (9)$$

Simulations on music signals where $\vec{x}_t$ were STFTs yielded good results, though abnormality early detection was not adequate in the context of music segmentation (where detection delay can be accepted).

4. SIMULATIONS

In this section, we compare our algorithm with two standard signal processing methods aimed at detecting abrupt changes, the Generalized Likelihood Ratio (GLR, see e.g. [1] for a detailed presentation) and a technique based on distance between time-frequency subimages [7]. As opposed to GLR, this technique is model-free. All the three techniques are tested both on artificial and real data.

The set of artificial data is composed of 2000 realisations of white gaussian noise of length 2048 filtered by an order 4 autoregressive model whose moduli are equal to 0.99, and frequencies are uniformely drawn between 0.05 and 0.45. In the first 1000 signals, the AR parameters have an abrupt change at point 1024, and the other 1000 are kept with no change. The three techniques are tuned as follows. As depicted in the previous sections, our algorithm is fed with two subsets of descriptors extracted from the smoothed pseudo Wigner-Ville of the input signal (time window of length 25, and frequency window of length 67). Each training vector is a TFR subimage of width 12, and the training set size is 20. The support vector parameters are: $\sigma = 1.5$ for the RBF gaussian kernel and $\nu = 0.2$. The same TFR is used for the distance-based method, with the TFR subimage width equal to 150, and we select the Kolmogorov distance [7]. Finally, the GLR is given the correct autoregressive model, with different orders: 3, 4 (correct order), and 6.

For a given input signal, each method yields an index whose maximum, if higher than a (method-dependent) threshold, corre-

sponds to an estimated abrupt change time instant. We define false alarms (FAs) as the detection of a change outside the neighborhood of the true change time instant, and true alarms (TAs) as the correct detection of an abrupt change inside that neighborhood. Fig. (3) plots the ROC curves (TAs vs. FAs) for the three methods. Both GLRs with autoregressive model orders 4 and 6 yield

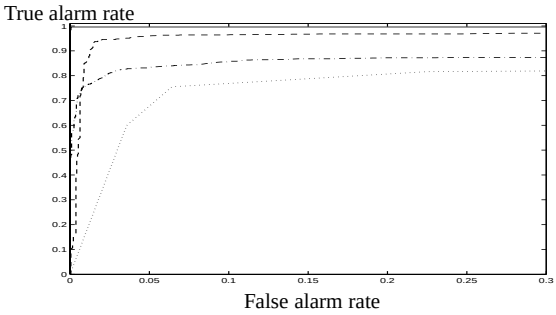

Fig. 3. ROC curves for the distance-based method (dot), GLR with a lower order (dash-dot), SVM method (dash), and GLR with correct or superior orders (solid). Very accurate performances are yielded by the GLR with correct (and superior order) and the SVM method.

very accurate results, which was expected as they were given the correct model. The SVM detector behaves accurately, with e.g. 95 per cent of true positives for 2 per cent of false positives, though no priors over the analyzed data are given. The TF distance-based method and the GLR with a lower order yield poor performances. Similar results can be observed on the histograms of the time instants corresponding to true positives (fig. 4). Note that the SVM

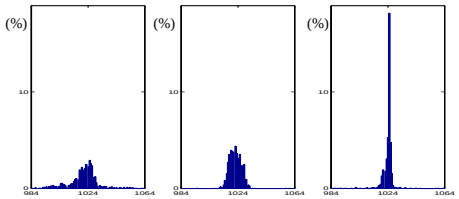

Fig. 4. Histograms of the estimated change time instant: percentage of the detected time instant inside the admissible neighborhood of the theoretic change time instant (1024) for the distance-based method (left), the SVM method (middle), and the GLR with correct order (right).

method gives better results than the TF distance-based technique though using the same descriptors as an input.

We have also tested our algorithm on music signals, where it proved to be quite efficient. Figure 5 displays the index obtained on the same music signal as in [6]. Every change of dynamics is properly detected (this is confirmed by listening the piece), which was not the case with the distance-based index (see [6]). Moreover, our index is much smoother, with abrupt changes corresponding to sharp peaks, and present high constrast on areas with and without changes.

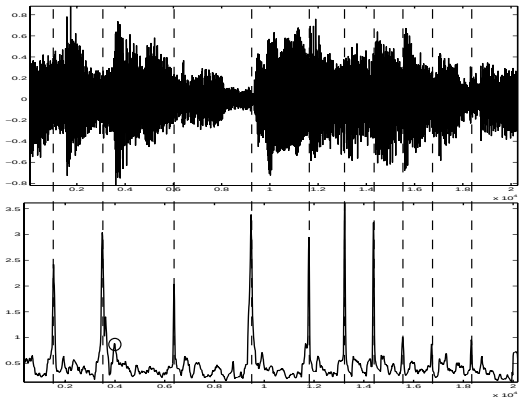

Fig. 5. Music signal (top) and the corresponding SVM abrupt change detection index (bottom). For a threshold η equal to 0.85, all changes (dashed lines) are correctly detected, with only one false positive (circled).

5. CONCLUSION

In this paper, we introduced an original kernel technique for detecting abrupt changes in signals. Simulations show its good behavior w.r.t. other algorithms on synthetic data, and superior performance on music signals, both in terms of accuracy of the detection and constrast.

6. REFERENCES

[1] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes - Theory and Application*, Prentice-Hall, April 1993.

[2] A. Smola and B. Schölkopf, *Learning with Kernels*, MIT press, 2002.

[3] B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, and R.C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[4] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, N.Y., 1995.

[5] A. Gretton and F. Desobry, "On-line one-class ν-support vector machines. an application to signal segmentation," in *IEEE ICASSP*, Hong-Kong, China, April 2003 (submitted).

[6] M. Davy and S. Godsill, "Detection of abrupt spectral changes using support vector machines. an application to audio signal segmentation," in *IEEE ICASSP-02*, Orlando, USA, May 2002.

[7] H. Laurent and C. Doncarli, "Stationarity index for abrupt changes detection in the time-frequency plane," *IEEE Signal Processing Letters*, vol. 5, no. 2, pp. 43 – 45, February 1998.