Time-Frequency Learning Machines For NonStationarity Detection Using Surrogates

Hassan Amoud, Paul Honeine, Cédric Richard, Pierre Borgnat, Patrick Flandrin

TIME-FREQUENCY LEARNING MACHINES FOR NONSTATIONARITY DETECTION USING SURROGATES

Hassan Amoud, Paul Honeine, Cédric Richard

Institut Charles Delaunay, FRE CNRS 2848
Université de technologie de Troyes
12 rue Marie Curie, 10010 Troyes France

Pierre Borgnat, Patrick Flandrin

Laboratoire de Physique, UMR CNRS 5672
École Normale Supérieure de Lyon
46 allée d'Italie, 69364 Lyon France

ABSTRACT

An operational framework has recently been developed for testing stationarity of any signal relatively to an observation scale. The originality is to extract time-frequency features from a set of stationarized surrogate signals, and to use them for defining the null hypothesis of stationarity. Our paper is a further contribution that explores a general framework embedding techniques from machine learning and time-frequency analysis, called time-frequency learning machines. Based on one-class support vector machines, our approach uses entire time-frequency representations and does not require arbitrary feature extraction. Its relevance is illustrated by simulation results, and spherical multidimensional scaling techniques to map data to a visible 3D space.

Index Terms— Time-frequency analysis, stationarity test, machine learning, one-class classification, surrogates.

1. INTRODUCTION

Time-frequency representations provide a powerful tool for nonstationary signal analysis and classification, supporting a wide range of applications [1]. As opposed to conventional Fourier analysis, these techniques reveal the evolution in time of the spectral content of signals. In [2, 3], time-frequency analysis is used to test stationarity of any signal. The proposed method consists of a comparison between global and local time-frequency features. The originality is to make use of a family of stationary surrogate signals for defining the null hypothesis of stationarity and, based upon this information, to derive statistical tests. An open question remains, however, about how to choose relevant time-frequency features.

Over the last decade, a number of new pattern recognition methods based on reproducing kernels have been introduced. These learning machines have gained popularity due to their conceptual simplicity and their outstanding performance [4]. Initiated by Vapnik's Support Vector Machines (SVM) [5], they offer now a wide class of supervised and unsupervised learning algorithms. In [6, 7, 8], the authors have shown how the most effective and innovative learning machines can be tuned to operate in the time-frequency domain. The present paper follows this line of research by taking advantage of learning machines to test and quantify stationarity. Based on one-class support vector machines, our approach uses the entire time-frequency representation and does not require arbitrary feature extraction. Applied to a set of surrogates, it provides the domain boundary that includes most of these stationarized signals. This allows us to test stationarity of the signal under investigation.

This paper is organized as follows. In Section 2, we introduce the surrogate data method to generate stationarized signals, namely, the null hypothesis of stationarity. The concept of time-frequency learning machines is presented in Section 3. In Section 4, it is applied to one-class SVM in order to derive a stationarity test. The relevance of the latter is illustrated by simulation results in Section 5 and 6.

2. STATIONARIZATION VIA SURROGATES

Stationarity refers to a strict invariance of statistical properties under every time shift. This theoretical definition can be loosely relaxed so as to encompass stationarity over some limited interval of observation. In order to test this property, it has been proposed in [2, 3] that a reference of stationarity be defined directly from the signal itself. The procedure consists of generating a family of stationarized signals which have the same energy density spectrum as the initial signal. Indeed, given an energy density spectrum, nonstationary signals differ from stationary ones by temporal structures encoded in the spectrum phase. The surrogate data technique [9] is an appropriate solution to generate a family of stationarized signals, since it destroys the time-varying structures in the signal phase while keeping its spectrum unchanged. In practice, this is achieved by keeping unchanged the magnitude of X, the Fourier transform of the initial signal x, and replacing its phase by a i.i.d. one. Each surrogate signal x_i results from the inverse Fourier transform of the modified spectrum, namely,

$$X_i(f) = |X(f)|\, e^{j\phi_i(f)},$$

This work was supported by ANR-07-BLAN-0191-01 STARAC.

978-1-4244-2710-9/09/$25.00 ©

with $\phi_i(f)$ drawn from the uniform distribution over the interval $[-\pi, \pi[$. This leads to as many stationary surrogate signals, $x_1, \ldots, x_n$, as phase randomizations $\phi_1(f), \ldots, \phi_n(f)$ are operated.

Once a collection of stationarized surrogate signals has been synthesized, different possibilities are offered to test the initial signal stationarity [2, 3]. A potential approach is to extract some features from the surrogates such as distance between local and global spectra, and to characterize the null hypothesis of stationarity by the statistical distribution of their variations in time. Another approach is based on statistical pattern recognition. It consists of considering surrogates as a learning set, and using it to estimate the support of the distribution of the stationarized signals. Stationarity of the signal under investigation can then be tested. Based on this concept, the method proposed in [2] relies on two features: fluctuations over time of the local power and mean frequency. The approach presented in this paper uses entire time-frequency representations, and does not require arbitrary feature extraction.

3. TIME-FREQUENCY LEARNING MACHINES

Most pattern recognition algorithms can be expressed in terms of inner products only, involving pairs of input data. Replacing these inner products with a (reproducing) kernel provides an efficient way to implicitly map the data into a high-dimensional space, and apply the original algorithm in this space. Calculations are then carried out without making direct reference to the nonlinear mapping applied to input data. This so-called *kernel trick* is the main idea behind (kernel) learning machines. In this section, we show that learning machines can be tuned to operate in the time-frequency domain by a proper choice of kernel. Refer to [8] for more details.

Time-frequency representations map a one-dimensional signal into the two-dimensional time-frequency domain in order to reveal its time-varying spectral content. Let $\mathcal{R}_x(t,f)$ denotes a given time-frequency representation of a signal x. Applying conventional pattern recognition algorithms directly to time-frequency representations usually consists of estimating a time-frequency pattern $\Psi(t,f)$ that maximizes a given criterion, based on a set of training signals denoted by $\{x_1, \ldots, x_n\}$. Examples of criteria include the maximum output variance for principal component analysis and the maximum margin for SVM. As illustrated in the next section with one-class SVM, these optimization problems only involve inner products between time-frequency representations of training signals:

$$\kappa(x_i, x_j) = \iint \mathcal{R}_{x_i}(t,f)\mathcal{R}_{x_j}(t,f)\,dt\,df = \langle \mathcal{R}_{x_i}, \mathcal{R}_{x_j} \rangle.$$

This can offer significant computational advantages. A well-known time-frequency representation is the spectrogram, defined by

$$S_x(t,f) = \left| \int x(\tau)\, h_{t,f}(\tau)\, d\tau \right|^2 = |\langle x, h_{t,f}\rangle|^2,$$

with $h_{t,f}(\tau) = h(\tau - t)\, e^{2j\pi f\tau}$. The inner product between two spectrograms, S_{x_i} and S_{x_j}, is given by the kernel [8]

$$\kappa_S(x_i, x_j) = \iint |\langle x_i, h_{t,f}\rangle\langle x_j, h_{t,f}\rangle|^2\, dt\, df.$$

Computing this kernel for any pair of surrogates yields

$$\kappa_S(x_i, x_j) = \iint \left| \langle |X|e^{j\phi_i}, H_{t,f}\rangle\, \langle |X|e^{j\phi_j}, H_{t,f}\rangle \right|^2 dt\, df, \tag{1}$$

where $H_{t,f}$ is the Fourier transform of $h_{t,f}$. For the sake of simplicity, we illustrate the proposed method with the spectrogram. However, one may easily extend this method to other time-frequency representations using their corresponding kernels, as given in [6, 8]. The Wigner distribution for instance, with its unitarity property, leads to substantial computational reduction since

$$\kappa_W(x_i, x_j) = |\langle x_i, x_j\rangle|^2 = \left| \int |X(f)|^2 e^{j(\phi_i(f) - \phi_j(f))}\, df \right|^2.$$

Any kernel learning machine can use time-frequency kernels to perform pattern recognition tasks in the time-frequency domain, as studied in [6, 7, 8]. In the next section, we present the one-class SVM problem to test stationarity.

We emphasize here that there is no need to compute and manipulate the surrogates and their time-frequency representations. Given $|X|$, only the random phases $\phi_i(f)$ and $\phi_j(f)$ are required to evaluate the kernel (1).

4. ONE-CLASS SVM FOR TESTING STATIONARITY

Inspired by SVM for classification, the one-class SVM allows the description of the density distribution of a single class [10]. The main purpose is to enclose the training data into a minimum volume hypersphere, thus defining a domain boundary. Any data outside this volume may be considered as an outlier, and its distance to the center of the hypersphere allows a measure of its novelty. Here, we propose to use the set of surrogate signals to derive the hypersphere of stationarity, in the time-frequency domain defined by a reproducing kernel as given in Section 3.

From a set of n surrogate signals, $x_1, \ldots, x_n$, computed from a given signal x, let $\mathcal{R}_{x_1}, \ldots, \mathcal{R}_{x_n}$ denote their time-frequency representations and κ the corresponding reproducing kernel. In this domain, we seek the hypersphere that contains most of these representations, with its center $\mathcal{C}$ and radius r given by solving the optimization problem

$$\min_{\mathcal{C}, r, \xi} \; r^2 + \frac{1}{n\nu}\sum_{i=1}^{n} \xi_i$$

$$\text{subject to } \; \|\mathcal{R}_{x_i} - \mathcal{C}\|^2 \le r^2 + \xi_i, \;\; \xi_i \ge 0, \;\; i = 1, \ldots, n,$$

with $\nu \in]0,1]$ a parameter controlling the tradeoff between minimizing the radius and controlling the number of training data outside the hypersphere, defined by the slack variables $\xi_i = (\|\mathcal{R}_{x_i} - \mathcal{C}\|^2 - r^2)_+$. Using the Lagrangian, the optimization problem is reduced to

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{n} \alpha_i \kappa(x_i, x_i) - \sum_{i,j=1}^{n} \alpha_i \alpha_j \kappa(x_i, x_j) \\ \text{subject to} \quad & \sum_{i=1}^{n} \alpha_i = 1, \;\; 0 \leq \alpha_i \leq \frac{1}{n\nu}, \;\; i = 1, \ldots, n, \end{aligned} \tag{2}$$

which can be solved with quadratic programming techniques. The resulting α_i's yield the center $\mathcal{C} = \sum_i \alpha_i \mathcal{R}_{x_i}$ and the radius $r = \|\mathcal{R}_{x_\ell} - \mathcal{C}\|$, with x_ℓ any data having $0 < \alpha_\ell < \frac{1}{n\nu}$.

For any signal x, the (squared) distance of its time-frequency representation from the center $\mathcal{C}$ can be written as

$$\|\mathcal{R}_x - \mathcal{C}\|^2 = \kappa(x,x) - 2\sum_{i=1}^{n} \alpha_i \kappa(x, x_i) + \sum_{i,j=1}^{n} \alpha_i \alpha_j \kappa(x_i, x_j).$$

As explained previously, we do not need to compute time-frequency representations to calculate this score, since only the values of the kernel are required. The coefficients α_i are obtained by solving (2), requiring only the evaluation of κ for training data. This *kernel trick* is also involved in the proposed decision function, defined by comparing the test statistics $\Theta(x) = \|\mathcal{R}_x - \mathcal{C}\|^2 - r^2$ to a threshold γ:

$$\Theta(x) \underset{\text{stat.}}{\overset{\text{nonstat.}}{\gtrless}} \gamma.$$

Thus, the initial signal x is considered as nonstationary if its time-frequency representation lies outside the hypersphere of squared radius $r^2 + \gamma$; otherwise, it is considered as stationary. Threshold γ has a direct influence upon the performance of the test [4]. For instance, with a probability greater than $1 - \delta$, one can bound the probability of false positive by

$$\Lambda = \frac{1}{\gamma n} \sum_{i=1}^{n} \xi_i + \frac{6R^2}{\gamma\sqrt{n}} + 3\sqrt{\frac{\ln(2/\delta)}{2n}}, \tag{3}$$

where R is the radius of the ball centered at the origin containing the support of the probability density function of the target class. Expression (3) can be used as a measure of stationarity of the signal x under investigation, see [3] for more details. Score Λ takes on a value of one if x is stationary and, the closer it is to zero, the greater the contrast between x and the surrogates is.

5. (SPHERICAL) MULTIDIMENSIONAL SCALING

Multidimensional scaling (MDS) is a classical tool in data analysis and visualization [11]. It aims at representing data in a lower-dimensional space, such that the resulting distances reflect, in some sense, the distances in the higher-dimensional space. The neighborhood between data is preserved, whereas dissimilar data tend to remain distant in the new space. MDS algorithm requires only the distances between data in order to embed them into the new space.

Consider the set of n time-frequency representations of the surrogate signals, $\{\mathcal{R}_{x_1}, \ldots, \mathcal{R}_{x_n}\}$, and the distances between each representations. We can apply classical MDS in order to visualize the n data in a 3D Euclidean space. The time-frequency domain has some properties that one can exploit here. For a normalized signal x, i.e., $\int |x(t)|^2 dt = 1$, the resulting surrogates and their time-frequency representations remain normalized, that is, $\kappa(x_i, x_i) = \langle \mathcal{R}_{x_i}, \mathcal{R}_{x_i} \rangle = 1$. This means that these representations lie on a unit-radius hypersphere centered at the origin. Thus it is more natural to embed them onto a sphere, as opposed to the classical Euclidean space. This can be performed by using a Spherical MDS technique, as proposed in [11], or more recently in [12]. The cosine of the *angle* between two time-frequency representations is defined by

$$\cos(\theta_{ij}) = \frac{\langle \mathcal{R}_{x_i}, \mathcal{R}_{x_j} \rangle}{\|\mathcal{R}_{x_i}\| \, \|\mathcal{R}_{x_j}\|} = \frac{\kappa(x_i, x_j)}{\sqrt{\kappa(x_i, x_i)\, \kappa(x_j, x_j)}}.$$

It is obvious that $\cos(\theta_{ij})$ is positive since $\kappa(x_i, x_j) \geq 0$. This gives us an information on the location of these representations on the hypersphere, since we thus have $-\frac{\pi}{2} \leq \theta_{ij} \leq \frac{\pi}{2}$.

6. EXPERIMENTATIONS

In order to test our method, we used the same two AM and FM signals as in [2]. While not covering all the situations of nonstationarity, these signals are believed to give meaningful examples. The AM signal is modeled as

$$x(t) = (1 + m \sin(2\pi t / T_0))\, e(t)$$

with $m \leq 1$, $e(t)$ a white Gaussian noise, T_0 the period of the AM. In the FM case,

$$x(t) = \sin 2\pi (f_0 t + m \sin(2\pi t / T_0))\, e(t)$$

with f_0 the central frequency. Based on the relative values of T_0 and the signal duration T, three cases can be distinguished for each type, AM and FM:

- $T \gg T_0$: The signal contains a great number of oscillations. This periodicity indicates a stationary regime.
- $T \approx T_0$: Only one oscillation is available. The signal is considered as nonstationary.
- $T \ll T_0$: With a small portion of a period, there is no change in the amplitude or the frequency. It is considered as a stationary signal.

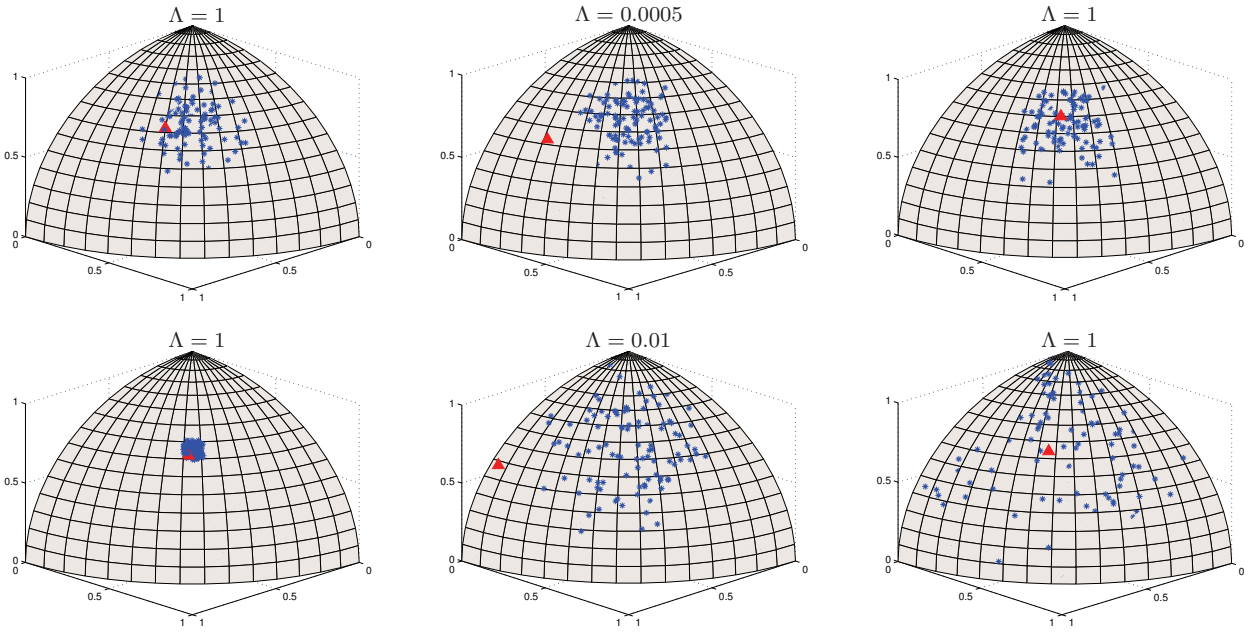

Fig. 1. Spherical 3D representation of the surrogates ($*$) and the tested signal ($\blacktriangle$), for the AM (first row) and the FM (second row), with $T \gg T_0$ (left), $T \approx T_0$ (middle) and $T \ll T_0$ (right).

For each signal, we considered $n = 100$ surrogate signals by generating their randomized phases to compute the corresponding spectrogram kernels (1). Using these kernel values, the one-class algorithm gave us the index of stationarity (3) for x. The Spherical MDS [12] was applied for each type, AM and FM, and each case, $T \gg T_0$, $T \approx T_0$ and $T \ll T_0$, as illustrated in Fig. 1. In the $T \approx T_0$ case, both AM and FM initial signals are outside the spread of the surrogates, thus indicating nonstationarity. In the other cases, $T \gg T_0$ and $T \ll T_0$, the corresponding signals are within the spread of the surrogates. These behaviors are endorsed by the one-class SVM results, with the score of stationarity given in Fig. 1. These results are consistent with those obtained in previous works, using either the distance or the time-frequency feature extraction approach [3]. Here, the test is performed without suffering from the prior knowledge required to extract relevant features.

7. CONCLUSION

In this paper, we showed how time-frequency kernel machines can be used to test stationarity. For a given signal, a set of stationarized surrogate signals is generated to train a one-class SVM, implicitly in the time-frequency domain by using an appropriate kernel. The latter is evaluated directly from the randomized phases, without computing time-frequency representations. The originality here is the use of the whole time-frequency information, as opposed to feature extraction techniques where prior knowledge is required. Its relevance has been illustrated with simulation results.

8. REFERENCES

[1] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, Academic Press Inc, 1998.

[2] J. Xiao, P. Borgnat, P. Flandrin, and C. Richard, "Testing stationarity with surrogates - a one-class SVM approach," in *Proc. IEEE Statistical Signal Processing*, Madison, WI, USA, August 2007, pp. 720–724.

[3] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing stationarity with surrogates: A time-frequency approach," *submitted to IEEE Trans. Signal Processing*, 2009.

[4] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.

[5] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.

[6] P. Honeine, C. Richard, P. Flandrin, and J-B. Pothin, "Optimal selection of time-frequency representations for signal classification: A kernel-target alignment approach," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, France, 2006, vol. 3, pp. 476–479.

[7] P. Honeine and C. Richard, "Signal-dependent time-frequency representations for classification using a radially gaussian kernel and the alignment criterion," in *Proc. IEEE Statistical Signal Processing*, Madison, WI, USA, August 2007, pp. 735–739.

[8] P. Honeine, C. Richard, and P. Flandrin, "Time-frequency learning machines," *IEEE Trans. Signal Processing*, vol. 55, pp. 3930–3936, July 2007.

[9] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1-4, pp. 77–94, 1992.

[10] D. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognition Letters*, vol. 20, no. 11-13, pp. 1191–1199, 1999.

[11] T. F. Cox and M. A. A. Cox, "Multidimensional scaling on a sphere," *Communications in Statistics: Theory and Methods*, vol. 20, pp. 2943–2953, 1991.

[12] R. Pless and I. Simon, "Embedding images in non-flat spaces," in *Proc. International Conference on Imaging Science, Systems and Technology*, H. R. Arabnia and Y. Mun, Eds., Las Vegas, 2002, pp. 182–188.