Signal Classification with an SVM-FFT Approach for Feature Extraction in Cognitive Radio

Manel Martínez Ramón
Departamento de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid
Avda de la Universidad, 30
28911 Leganés, Madrid, Spain
Email: manel@ieee.org

Thomas Atwood
Department of Electrical and Computing Engineering
The university of New Mexico
1, University of New Mexico
87131-001, NM, USA
Email: tatwood@eece.unm.edu

Silvio Barbin
University of São Paulo
São Paulo, SP - Brazil
Email: barbin@usp.br
Center for Information Technology Renato Archer
Campinas, SP Brazil
silvio.barbin@cti.gov.br

Christos G. Christodoulou
Department of Electrical and Computing Engineering
The university of New Mexico
1, University of New Mexico
87131-001, NM, USA
Email: christos@eece.unm.edu

***Abstract*—The estimation of the spectrum usage from the point of view of number of users and modulation types is addressed in this paper. The techniques used here are based on Support Vector Machines (SVM). SVMs are machine learning strategies which use a robust cost function alternative to the widely used Least Squares function and that apply a regularization which provides control of the complexity of the resulting estimators. As a result, estimators are robust against interferences and nongaussian noise and present excellent generalization properties where the number of data available for the estimation is small. The structure presented here has a feature extraction part that, instead of using an FFT approach, uses the SVM criterion for spectrum estimation, feature extraction and modulation classification.**

I. Introduction

Cognitive Radio (CR) is an emerging technology that promises to dramatically increase the utilization of the available radio resources, as well as to dramatically change the way in which a user interfaces with a communication device. Following [1] the key features of a CR are 'the awareness of the radio environment in terms of spectrum usage, power spectral density of transmitted/received signals, wireless protocol) and intelligence." A CR can be thought as a software defined radio (SDR) with possibly reconfigurable antennas that provide flexibility and reconfigurability plus a machine learning device (MLD) that provides the needed intelligence to adapt the SDR to the given environment through a set of trade-offs between some optimallity criteria and some user, system or environment constraints (Figure 1) [2]. In particular, the MLD is intended to learn how the users interact with the radio, and how best to use the available radio resources.

A lot of the MLD research has focused on Genetic Algorithms or Neural Networks for both of these tasks [3], [4]. An interesting approach to characterize the users is to determine the types of modulations used. Hu and coworkers [5] proposed a strategy to classify among different modulations by using a feature extraction method based on spectral correlation analysis [6], [7]. The extracted features were fed into a multiclass Support Vector Machine (SVM) [8], [9] classifier. Hu's method showed excellent results in gaussian noise environment. Authors compared among different classification algorithms to state that the SVM approach is the best one when the number of data available for the estimation is small. Hu's method is ultimately based on the use of the Fast Fourier Transform for spectrum estimation and feature extraction.

Spectrum estimation has to be performed using small, broadband and nondirective antennas with the assumption of heavy noise and interference in the RF channel. When the interferences are not Gaussian, methods that rely on a Least Squares (LS) optimization criterion may not produce accurate spectrum estimation. Among those methods, the most widely used is the Discrete Fourier Transform (DFT). The Minimum Variance Distortionless Response method (MVDR) and MUltiple SIgnal Classification (MUSIC) are also widespread methods that implicitly use the LS criterion, thus lacking accuracy in non Gaussian scenarios.

In [10] a method for spectrum sensing has benn presented based on the SVM method ontroduced in [11]. This method implicitly uses a cost function which is linear, thus being part of so-called robust regression methods. Moreover, SVM has been proven to use a cost function which has the same properties as the Huber Robust Cost function [12]. While these methods are suboptimal under Gaussian noise, they are very robust under non Gaussian noise conditions.

Here we present a methodology that combines both strate-

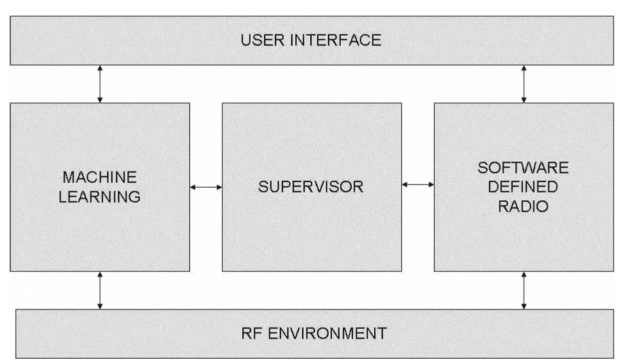

Fig. 1. Cognitive radio block diagram.

978-1-4244-5357-3/09/$26.00©2009IEEE

gies and that produces an estimator that holds the features of the SVM in both the feature extraction part and the classification part.

In the next section the techniques used in [11] and in [10] for spectrum estimation are described, and in section III the feature extraction used for signal classification in conbination with the spectrum SVM estimation is presented. Section IV presents the results of this combination in some modulation classification experiments.

II. Spectrum Estimation based on Support Vector Machines

Support Vector Machines (SVM) are a class of learning machines whose criterion for optimization consists of a trade-off between the minimization of the training error and the minimization of the quadratic norm of the parameter vector. This last term is a regularization that controls the generalization ability of the machine, thus improving the performance with respect to non-regularized methods. For a regression model of the form $y[i] = \mathbf{w}^T\mathbf{x}[i]$, the functional that includes both terms is

$$L = \frac{1}{2}\left\|\mathbf{w}\right\|^2 + C\sum_{i=1}^{N}\ell\left(\xi_i + \xi_i'\right) \tag{1}$$

subject to

$$\begin{aligned} \mathbf{w}^T\mathbf{x}[i] - y[i] &= \xi_i + \varepsilon \\ -\mathbf{w}^T\mathbf{x}[i] + y[i] &= \xi_i' + \varepsilon \\ \xi_i, \xi_i' &\geq 0 \end{aligned} \tag{2}$$

where $\ell(\cdot)$ is a convex cost function, C is the trade off parameter and ξ_i, ξ_i' are the slack variables or losses. The constraints mean that, provided the slack variables must be positive or zero, if the error is between $\pm\varepsilon$, then this error is not taken into account. Otherwise, for positive or negative errors, we minimize the contribution of the slack variables.

Applying Lagrange multipliers α_i, α_i' to each constraint of (2) on functional (1) leads to the following equivalent functional

$$-\frac{1}{2}\left(\alpha - \alpha'\right)^T\mathbf{K}\left(\alpha - \alpha'\right) + \left(\alpha - \alpha'\right)^T\mathbf{y} - \varepsilon\mathbf{1}^T(\alpha - \alpha') \tag{3}$$

with

$$\begin{aligned} \mathbf{w} &= \mathbf{X}\left(\alpha - \alpha'\right) \\ \mathbf{K} &= \mathbf{X}^T\mathbf{X} \end{aligned} \tag{4}$$

where $\boldsymbol{\alpha}$ is a vector containing all Lagrange multipliers, and $\mathbf{X}$ contains all training input vectors in column form. Functional (4) is usually solved through quadratic programming.

In spectrum estimation, we assume a linear model [11]

$$y[n] = \sum_{k=1}^{K} a_k \cos\left(\frac{2k\pi}{K}n\right) + b_k \sin\left(\frac{2k\pi}{K}n\right) \tag{5}$$

so identifying terms

$$\begin{aligned} \mathbf{w} &= \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \\ \mathbf{x}[n] &= \left(1, \cdots \cos\left(\tfrac{2k\pi}{K}n\right), \cdots, 1, \cdots \sin\left(\tfrac{2k\pi}{K}n\right)\cdots\right)^T \end{aligned} \tag{6}$$

In order to compute the spectrum, we simply solve functional (4) and then compute terms a and b using (6). Straightforwardly, the spectrum estimation at frequency $\omega_k = \frac{2k\pi}{K}$ is

$$Y(k) = \left\|a_k + jb_k\right\|^2 \tag{7}$$

This spectrum is similar to the DFT spectrum, but here we do not assume that the noise is white. Instead, we use a cost function that is zero between $\pm\varepsilon$, and that it is linear beyond $\pm\varepsilon \pm \gamma C$. This gives more robustness against non Gaussian interferences. Also, the regularization parameter improves the generalization with respect to the one of the quadratic criterion implicit in DFT.

Note that instead of using sinusoids as approximating functions, one may use ad-hoc signals as modulated pulses to better approach the spectrum where a priori knowledge about the signal under detection is available [13].

III. Feature extraction based on spectral correlation

Ciclostationarity properties of modulated signals were first derived by Gardner in the middle 80's [6], [7], and they are commonly used to extract time and frequency domain features that are used to classify the signals among a given set possible modulations. Time domain analysis comprises the cyclic autocorrelation function. For a deterministic time series $x(t)$, we define the cyclic autocorrelation function as

$$\hat{R}_y^\alpha(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} y(u+\tau/2)y^*(u-\tau/2)e^{-i2\pi\alpha u}\,du \tag{8}$$

The series is said to be wide-sense cyclostationary with period T_0 if $\hat{R}_x^\alpha$ is not identically zero for $\alpha = nT_0$ for some integers n, but is identically zero for all other values of α.

The Fourier transform of this signal is the so-called spectral correlation function, and it can be expreseed as

$$S_y^\alpha(f) = \lim_{T\to\infty} S_{yT}^\alpha(t, f) \tag{9}$$

where $S_{xT}^\alpha(t, f)$ is the Fourier transform of the expression after the limit in equation (8), this is,

$$S_{xT}^\alpha = \frac{1}{T}Y(t, f + \frac{\alpha}{2}) \cdot Y^*(t, f - \frac{\alpha}{2}) \tag{10}$$

where $Y(t, f)$ is the Fourier transfor of the signal $y(t)$ in the interval $t \pm T/2$. In our approach, we substitute the Fourier transform by the SVM-DFT transform of the previous section.

The spectral correlation function can easily be particularized to a finite interval in order to make it usable in practice. In [5], authors also use the so called spectral autocorrelation coefficient between frequency components placed at a distance $\pm\alpha/2$, which is expressed as

$$C_y^\alpha(f) = \frac{S_y^\alpha(f)}{\sqrt{(S_y^0(f+\alpha/2)S_y^0(f-\alpha/2))}} \tag{11}$$

Four characteristics are used in [5] and other works to classify modulations using machine learning algorithms. The first one consists of the count of narrow pulses in the frequency domain present in the spectral autocorrelation function. To find this feature, it is enough to set $\alpha = 0$ and count the number of peaks in the resulting function. The second feature to extract is the number of spectral lines in the α domain of the spectral autocorrelation function. The third feature is the average energy of these pulses. The fourth feature of the set is the maximum value of the spectral correlation coefficient.

IV. Results

Several experiments have been run in order to compare the results of the signal classification using the SVM-FFT approach, and we compared them with those of a standard FFT spectrum in a strong non Gaussian interference environment.

A. Experiment setup

The structure of our algorithm consists of three stages:

- An SVM-DFT estimator. This stage takes the signal and, using the formulation of section II, computes an estimation of the spectrum of the signal. Note that this SVM does not need a training and test parts. Instead, the desired result is the set of coefficients a_i and b_i in eq. (6). To do that, we optimize eq. (3) and compute the coefficients using eq. (4). This spectrum is used to compute the spectral correlation function and the spectral correlation coefficients.
- A feature extractor, which extracts the four features described in section III from the spectral analysis performed by the previous stage.
- A standard SVM classifier used to classify among the various modulations. The binary version of the SVM is completely described in [9]. Nevertheless, we need, in general, a multiclass classifier in order to be able to include more than two class of modlulations. There are several approaches that can be constructed using a binary base classifier, as the one-against-all, the one-against-one, output correcting codes [14] or the directed acyclic graphs [15], among others. We use the direct multiclass SVM [16], which is implemented with the software LIB-SVM [17]. The kernel function chosen for the SVM is the Gaussian radial basis function, as they have largely demonstrated excellent approximation properties in a variety of classification and regression applications.

In this research we are using MATLAB/SIMULINK to generate a representation of several modulations. We simulate an ISM-like environment with signals that travel along a deep fading Raileigh channel, simulating the common multipath environment of indoor signals. Signals are sampled at Nyquist frequency in order to obtain their complex envelope equivalents. The signal has been corrupted with AWGN and with impulse noise. Impulse noise is generated by a train of impulses distributed in time with a Bernouilly probability density, and with probability of occurrence of 10%, whose amplitude had a Gaussian statistics. The data used to train the classifier was corrupted only with Gaussian noise of

No of symbols	1	2
SVM	-0.6 $\pm$ 2.00	-0.05 $\pm$ 0.10
FFT	8.22 $\pm$ 16.34	2.00 $\pm$ 11.22
No of symbols	3	4
SVM	-0.01 $\pm$ 0.22	-0.00 $\pm$ 0.00
FFT	4.09 $\pm$ 10.09	1.91 $\pm$ 3.32

TABLE I
FREQUENCY ESTIMATION OF SPECTRAL PEAKS USING SVM-FFT AND STANDARD FFT.

variance 0.1, and consisted of several thousands of patterns. The training of the structure was performed online, and a cross validation of parameters C and σ of the SVM was run using the common v-fold technique.

The test signal was corrupted by Gaussian noise with impulse noise added in blocks of random duration and length. These blocks of noise have a probability of occurrence of 1%, and they simulate an interference. The Gaussian noise power was 0 dB and the block noise power is variable.

There are four classes of signal modulations: BPSK, QPSK, FSK and MSK, but more modulations can be considered, as general QAM signals.

B. Robustness against interferences

Usually, a classifier that has been trained with a set of signals corrupted with a fixed signal to noise ratio, will present poorer performance with test signals of different SNR, when compared with a classifier trained and tested with the same SNR. This is due to the fact that the optimal classification boundary changes with the SNR, because the statistics of the features also change. This problem can be partly alleviated if the feature extractor is robust against the interference, this is, if for a given range of noise power, the features do not significantly change. This is one of the features of the SVM-FFT. In order to compare the robustness of the classifiers against interferences, the test signal has been corrupted with different values of impulse noise power, and the classifier tested with them.

In order to show the performance of the SVM-FFT against the standard FFT, we performa a peak detection test over BPSK signals in impulse noise for different number of symbols. Table I shows the mean and variance of the resulting estimations. As it can be seen, the SVM-FFT is able to detect the spectral peaks with small error where, in some cases, the FFT detection error is not acceptable. This result is consistent with the one presented in [10].

Table II shows the performance in the classification of the signals with no impulse noise, and Gaussian noise with $SNR = 3dB$. Table III shows the classification accuracy with 7 dB impulse noise power. Where both approaches have similar performance in conditions of Gaussian additive noise, it can be seen that the classification error of the feature extractor based on the FFT significantly degrades in presence of impulse noise, not considered during the training, while the SVM-DFT has a reasonable good performance.

FFT	BPSK	QPSK	FSK	MSK
BPSK	99.3	0	0.4	0.5
QPSK	0	100	0	0
FSK	0.7	0	99.6	0
MSK	0	0	0	99.5
SVM-FFT	BPSK	QPSK	FSK	MSK
BPSK	99.5	0	0	0
QPSK	0	100	0.3	0
FSK	0.5	0	99.7	0
MSK	0	0	0	100

TABLE II
CLASSIFICATION ACCURACY FOR THE SVM WITH FFT-BASED FEATURE EXTRACTOR AND FOR THE SVM WITH SVM-DFT-BASED EXTRACTOR. NO IMPULSE NOISE ADDED.

FFT	BPSK	QPSK	FSK	MSK
BPSK	59.3	15.7	33.4	21.5
QPSK	20.4	50.7	12.2	12.7
FSK	13.5	11.9	49.9	16.3
MSK	6.5	21.7	4.5	49.5
SVM-FFT	BPSK	QPSK	FSK	MSK
BPSK	96	1	1.1	1.3
QPSK	0	99	0.3	2.1
FSK	3.5	0	89.2	0.6
MSK	0.5	0	9.4	96

TABLE III
CLASSIFICATION ERROR FOR THE SVM WITH FFT-BASED FEATURE EXTRACTOR AND FOR THE SVM WITH SVM-DFT-BASED EXTRACTOR. IMPULSE NOISE ADDED.

V. CONCLUSION

Spectrum sensing is a key tool for CR. This task can challenging due to noise and interferences in many possible scenarios. We particularly point out the situation where small low gain antennas are used in presence of heavy non Gaussian interferences. We introduced here a method based on Support Vector Machines that are robust against heavy interferences, in scenarios in which DFT is unable to estimate the spectrum due to the fact that the statistics of the noise is non Gaussian.

We briefly presented the standard SVM technique, and its application to spectrum estimation. For this purpose we used an algorithm that has the same structure as the DFT, but removing the LS criterion of optimality. Instead, we use the SVM criterion, which involves a robust cost function and a regularization term that limits the complexity of the resulting estimator. We propose here its use to detect modulations in noise, using this technique as a feature straction step in a standard spectral correlation estimation presented in [5].

In the simulations we show that when the signal is corrupted by heavy interference, SVM produces a good estimate while the FFT does not. The simulations show that the SVM has a better repeatability and accuracy, and the differences in performance increase with the interference power.

Finally, it is important to remark that, as FFT is optimal in a Gaussian noise environment, it shows the same performance as the SVM technique in these conditions. In other words, SVM has no advantage over the DFT under Gaussian noise, as LS is the optimal criterion in this scenario.

REFERENCES

[1] E. Hossain and V. K. Bhargava, Eds., *Cognitive Wireless Communication Networks*. NY: Springer, 2007.

[2] B. Fette, *Cognitive Radio Technology*. New York, NY: Elsevier, 2006.

[3] N. Baldo and M. Zorzi, "Learning and Adaptation in Cognitive Radios using Neural Networks," in *Proceedings of the IEEE CCNC 2008*, 2008, pp. 998–1003.

[4] Z. Zhang and X. Xie, "Intelligent Cognitive Radio: Research on Learning and Evaluation of CR Based on Neural Network," in *Proceedings of the 5th International Conference on Information and Communications Technology*, 2007, pp. 33–37.

[5] H. Hu, J. Song, and Y. Wang, "Signal Classification based on Spectral Correlation Analysis and SVM in Cognitive Radio," in *Proceedings of the 22nd International Conference on Advanced Information Networking and Applications (AINA)*, 2008, pp. 883–887.

[6] W. Gardner, "Spectral Correlation of Modulated Signals: Part I–Analog Modulation," *Communications, IEEE Transactions on*, vol. 35, no. 6, pp. 584–594, Jun 1987.

[7] W. Gardner, W. Brown, and C.-K. Chen, "Spectral Correlation of Modulated Signals: Part II–Digital Modulation," *Communications, IEEE Transactions on*, vol. 35, no. 6, pp. 595–601, Jun 1987.

[8] A. J. Smola and S. B, "A Tutorial on Support Vector Regression," *Statistics and Computing*, vol. 4, no. 3, pp. 199–222, Aug. 2004.

[9] C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 1–32, 1998.

[10] T. D. Atwood, M. Martnez-Ramon, and C. G. Christodoulou, "Robust Support Vector Machine Spectrum Estimation in Cognitive Radio," in *Proceedings of the 2009 IEEE International Symposium on Antennas and Propagation and USNC/URSI National Radio Science Meeting*, 2009.

[11] J. L. Rojo-Álvarez, M. Martínez-Ramón, A. R. Figueiras-Vidal, A. García-Armada, and A. Artés-Rodríguez, "A robust support vector algorithm for nonparametric spectral analysis," *IEEE Signal Processing Letters*, vol. 10, no. 11, pp. 320–323, Nov 2003.

[12] K.-R. Müller, A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, "Predicting time series with support vector machines," in *Advances in Kernel Methods — Support Vector Learning*, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge, MA: MIT Press, 1999, pp. 243–254.

[13] T. D. Atwood, M. Martnez-Ramon, and C. G. Christodoulou, "Parametric Spectral Occupancy Estimation Using Signal Autocorrelation Functions," in *Proceedings of the 2009 IEEE International Symposium on Antennas and Propagation and USNC/URSI National Radio Science Meeting*, 2009.

[14] T. G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error correcting output codes," *J. Artificial Intelligence Res.*, vol. 2, pp. 263–286, 1995.

[15] J. Platt, N. Cristianini, and J. Shawe-Taylor, *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2000, vol. 12, ch. Large Margin DAG's for Multiclass Classification, pp. 547–553.

[16] J. Weston and C. Watkins, "Multi-Class Support Vector Machines," in *Proc. ESANN99*, M. Verleysen, Ed., Brussels, Belgium, 1999.

[17] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.