ROBUST SIGNAL CLASSIFICATION USING THE WAVELET TRANSFORM FOR FEATURE EXTRACTION

Frank Brickle, PhD
ACPT, Inc.
3 Nami Lane, Unit C-9
Hamilton, NJ 08619
brickle@pobox.com

ABSTRACT

The success of a classification method will depend largely on the *independence* and *expressivity* of the features it chooses to observe. In that regard the Discrete Wavelet Transform ought to be especially useful, thanks to its orthogonality properties. We have already developed applications which use Wavelet Transforms of Power Spectrum estimates as the basic observations for classifying a variety of digital signals in adverse HF environments. Even with the simplest of scoring and modeling techniques, these applications have proved to be extremely accurate, robust, and efficient. However, these applications represent only a first step into a more systematic exploitation of Wavelet-based observations for general signal classification. Starting with an informal illustration of the original applications, we describe ongoing work at refining and extending Wavelet techniques towards a comprehensive system for signal classification. In particular we address issues in multiway classification, markovity versus cyclostationarity, online training and updating, and heuristic methods for reducing the computational overhead associated with complex Wavelets and quadrature signals.

1. INTRODUCTION

In this paper we describe a novel general method for automatic classification of a large class of waveforms. The technique described has been implemented and used succesfully in a variety of fielded applications. It is simple, efficient, very accurate, and robust.

The success of a classification method will depend largely on the *independence* and *expressivity* of the features it chooses to observe. In that regard the Discrete Wavelet Transform [1] is found to be especially useful, thanks to its orthogonality properties.

The first key element of the technique is the extraction of features from power spectrum estimates using the Wavelet Transform. In this sense, the technique constitutes a form of signature or template matching against an exemplar derived from live data. However the second key feature is the accumulation of evidence over a sequence of trial matches. It is this accumulation of evidence that provides much of the discrimination power in adverse signal environments. The third key feature is an array of final tests that are applied only when the weight of accumulated evidence exceeds certain predetermined thresholds. The robustness of the technique is afforded in large measure by these additional tests.

This method has proved effective in applications where either the aggregate signal environment is adverse (atmospheric or impulsive noise, adjacent channel interference, selective fading) or the collection point is disadvantaged (poor SNR, receiver mistuning).

Our aim in this discussion is to provide an informal overview of the method rather than a rigorous exposition of the component steps. With that as a starting point, we proceed to outline the directions in which the technique can be expanded and refined.

2. THE BASIC OBSERVATIONS

The basic observations in our technique are sequences of power spectrum computations. An ongoing signal is transformed into a series of data frames representing power spectrum estimates computed by FFT. Each frame is then normalized to a peak power of 0dB, clipped at the low end to a uniform minimum (typically -48dB), and justified in frequency such that the total frequency span represented is the same for all signals (typically 4kHz represented by 1024 bins). Each processed frame is then subjected to the Forward Discrete Wavelet Transform (typically using a Daubechies 20 [2] mother wavelet). From the resulting frame of wavelet coefficients, a contiguous subset of coefficients is extracted (typically, the low ½ of the points). In addition the Wavelet Power Spectrum (WPS) [3] is computed from the full set of coefficients.

The resulting subvector of coefficients, along with normalized WPS values, constitute the feature set of a single frame of the input signal.

The chief difference between observation sequences and exemplars is merely that an exemplar is represented by a single subvector-and-WPS collection, which was pre-computed from the total estimated power spectrum of a segment of live data of interest. In a later section we will touch on the uses of multiple exemplar representations, as they may apply to signals exhibiting significant markovity.

3. COMPARISON AND SCORING

The main iteration in the technique consists of repeatedly forming the next observation frame from the candidate signal, and comparing it with the exemplar frame. Each comparison yields a score. The score for each frame is published to a supervisory procedure.

There are a number of ways to compare the frames, amounting to score computations that are distributed statistically differently for different applications. For purposes of this discussion, it suffices that the comparison yield some estimate of the distance [4] between the observation and exemplar wavelet coefficient vectors.

For example, one comparison might be performed by computing the vector cosine between the wavelet coefficient subvectors of the observation and exemplar frames.

Roughly, "close" vectors are understood to correspond to similar spectra. It should be easy to see that accumulation of scores is needed since the variance of estimates in short-term spectral frames. A comparison between observations made on overall spectral estimates would be effective, but since the desire is for an online system, the method of accumulating scores makes it possible to identify matches as close as possible to their first occurrence in the candidate signal.

An additional twist is applied in the "real" implementation. The frame-by-frame comparison is applied in fact to segments of the power spectrum vectors at multiple offsets, and the best match distance and offset are published. This adds little complexity to the computation but eliminates most of the effects of receiver mistuning.

4. PRIMARY SCORE EVALUATION

As scores are produced by the comparison operation, they are published to a supervisory process. The job of the supervisor is simple. It merely takes the most recent score and adds it to the current aggregate score for the candidate stream. If the resulting aggregate score exceeds a predetermined threshold, the supervisor proceeds to execute secondary tests.

5. SECONDARY SCORE EVALUATION

The Primary test threshold is tunable, in the sense that it can be tweaked to be more or less permissive. (It will be seen that in the rigorous version of this method, favoring Type I or Type II errors can have significant consequences for multiway classification.) A more permissive setting puts a great deal of emphasis on secondary testing. In this technique, the secondary test consists of a similarity measure much like the main spectrum/wavelet coefficient vector distance, but carried out on the WPS. The test consists of a single scalar comparison of distances between the WPS values of the observation stream to date and the exemplar. For wavelet-based features, the WPS is effective at summarizing the general *concentration* of details. In other words, it is capable of distinguishing whether similar features also happen in the "right place" in the spectrum. This is especially important in discriminating among closely-related signals.

Observation streams that pass the secondary test are signaled as showing onset of the signal of interest.

In some applications, the entire process is continued indefinitely, since the disappearance of the signal of interest may also be consequential.

6. HOW GOOD IS IT?

One realization of this method has been integrated into a system for classifying a family of HF signals. For this application, a complete suite of exemplars was prepared from live sample data covering the total repertoire of signals emitted by devices from a single commercial vendor. The goal was to distinguish these signals from a fairly extensive universe of signals of every kind found on HF.

The classifiers associated with these exemplars were subjected to an extensive battery of tests against over 10,000 signal samples of virtually every common type, down to 3dB SNR and up to 20dB channel fading, with up to 3kHz mistuning. A typical result was 100% success at identifying signals of interest within 200ms of onset, with a false alarm rate of less than 7%. Typically, a classifier algorithm would run at 16 times real time on a 1.5GHz consumer-level desktop.

On another front, an early version of this technique is employed by the VAD detection system running on the Agilent E3238S Signal and Intercept Collection Station.

7. EXTENSIONS AND REFINEMENTS

7.1. Multiway classification

The method described thus far only applies to distinguishing a single exemplar from all other possibilities. It is desirable to extend the method to distinguish one among a *set* of possibilities, from all remaining possibilities. The most favorable form of this configuration would be a process that assigned weights for a given signal as an instance of any of the items of interest.

The method of accumulating evidence from comparison with exemplars creates difficulties here. The chief reason is that, at root, the technique measures *non-normalized distances* between vectors. It is not meaningful to compare scores between different exemplars. The upshot is that accumulating sequences of evidence may reach their determining thresholds at different stages.

If we want classifiers to run in precise parallel, the comparisons must in fact be converted to evaluations of likelihood based on probability models. The essential component of such a change is the ability to normalize models based on *misclassification cost* [5]. This subject will be treated in greater depth in a subsequent paper.

7.2. Markovity and Cyclostationarity

A perceived liability of this technique is its apparent dependence on stationarity in the candidate signals. Cyclostationarity [6] appears to circumvent this shortcoming. However, we conjecture that cyclostationarity is a sub-case of the larger issue of markovity among probability models. This, too, we will address in a forthcmoing discussion.

7.3. Real Wavelets with Complex Signals

Another shortcoming of this technique is the fact that it is confined to real (that is, not quadrature) signals. Furthermore complex wavelets are usually seen as incurring disproportionately high computational overhead. We have developed a technique, however, for reducing wavelet computations on specifically quadrature signals, to real wavelet computations. This technique relies on the fact that all frequency components in quadrature signals exhibit constant phase offset, which survives any modification of the magnitude components of the complex data. Once again, this algorithm will be detailed in a forthcoming paper.

8. CONCLUSION

We have described a simple and effective method for applying the Discrete Wavelet Transform to sequences of power spectrum estimates, for the purpose of automatically classifying signals of interest. This method has been realized and successfully applied in a number of fielded systems. Extensions and refinements of the method to cover a wider variety of signal classes and presentations are straightforward and under development.

9. REFERENCES

The references in the text should be numbered in order of appearance. When referring to them in the text, type the corresponding reference number in square brackets, as shown at the end of this sentence [1]. If the same reference is used more than once in the text, refer to it in later text by the first reference number. List all bibliographical references by consecutive numbers at the end of the paper in Times Roman 9 pt. type with 10 pt. line spacing. Note that the references in the list hang indent by .25 inch (6 mm) and there is no line spacing between them. Use authors' initials and last names, and cite the article or book according to the sample references shown below.

[1] Mallat, Stephen, *A Wavelet Tour of Signal Processing,* Elsevier, Burlington, MA 1999.

[2] Daubechies, Ingrid, *Ten Lectures on Wavelets*, SIAM 1992.

[3] Jense, A., laCour-Harbo, A., *Ripples in Mathematics: the Discrete Wavelet Transform*, Springer, 2001

[4] Apostol, T., *Calculus, Vol. 1: One-Variable Calculus with an Introduction to Linear Algebra,* John Wiley and Sons 1967.

[5] Beritelli.F; Casale.S; Ruggeri.G; Serrano.S, "Performance evaluation and comparison of G.729/AMR/fuzzy voice activity detectors", Signal Processing Letters, IEEE ,Vol. 9 , Issue 3 , March 2002, pp.85 - 88

[6] Gardner, William A.; Antonio Napolitano, and Luigi Paura, "Cyclostationarity: Half a century of research". *Signal Processing* **86** (4): 639–697. Elsevier 2006.

