CHIN.PHYS.LETT. Vol. 23, No. 7 (2006) 1951

Wavelet Space Partitioning for Symbolic Time Series Analysis *

Venkatesh Rajagopalan, Asok Ray**

Department of Mechanical and Nuclear Engineering, College of Engineering, The Pennsylvania State University, University Park, PA 16802-1412, USA

(Received 4 March 2006)

A crucial step in symbolic time series analysis (STSA) of observed data is symbol sequence generation that relies on partitioning the phase-space of the underlying dynamical system. We present a novel partitioning method, called wavelet-space (WS) partitioning, as an alternative to symbolic false nearest neighbour (SFNN) partitioning. While the WS and SFNN partitioning methods have been demonstrated to yield comparable performance for anomaly detection on laboratory apparatuses, computation of WS partitioning is several orders of magnitude faster than that of the SFNN partitioning.

PACS: 89.75.−k, 89.70.+c, 07.90.+c

Symbolic time series analysis (STSA) has been proposed for real-time anomaly detection in complex systems.[1−3] A crucial step in STSA is partitioning of the phase space of the underlying dynamical system for symbol sequence generation.[4] Several techniques have been suggested in the physics literature for symbol generation, primarily based on phase space partitioning.[5,6]

Symbolic false nearest neighbour (SFNN) partitioning[7] optimizes a generating partition by avoiding topological degeneracies. The criterion is that short sequences of consecutive symbols ought to localize the corresponding state space point as closely as possible. This is achieved by forming a particular geometrical embedding of the symbolic sequence under the candidate partition and minimizing the apparent errors in localizing the state space points. In a good partition, nearby points in the embedding remain close when mapped back into the state space. In contrast, bad partitions induce topological degeneracies where symbolic words map back to globally distinct regions of state space. The nearest neighbour to each point in the embedding is found in terms of Euclidean distance between symbolic neighbours, where better partitions yield a smaller proportion of symbolic false nearest neighbours. For convenience of implementation, the partitions are parameterized with a relatively small number of free parameters. This is accomplished by defining the partitions with respect to a set of radial-basis 'influence' functions. The statistic for symbolic false nearest neighbours is minimized over the free parameters using 'differential evolution', a genetic algorithm suitable for continuous parameter spaces.[7]

A major shortcoming of SFNN partitioning is that it may become extremely computation-intensive if the dimension of the phase space of the underlying dynamical system is large. Furthermore, if the time series becomes noise-corrupted, then the symbolic false neighbours rapidly grow in number and require a large symbol alphabet to capture the pertinent information on the system dynamics. The wavelet transform largely alleviates these shortcomings and is particularly effective with noisy data from high-dimensional dynamical systems. Gamero *et al.*[8] used wavelet-based multiresolution for signal analysis in time-dependent nonlinear systems.

In this Letter, we present a novel concept of wavelet-space (WS) partitioning that relies on entropy maximization. The WS and SFNN partitioning methods are compared from the perspectives of anomaly detection. The results are obtained based on time series data generated from two laboratory apparatuses: (i) a nonlinear electronic system, and (ii) a mechanical vibration system.

In multi-resolution analysis (MRA) of wavelet transform, a continuous signal $f \in \mathbb{H}$, where $\mathbb{H}$ is a Hilbert space, is decomposed as a linear combination of time translations of scaled versions of a suitably chosen scaling function $\phi(t)$ and the derived wavelet function $\psi(t)$. Let the sequence $\{\phi_{j,k}\}$ belong to another Hilbert space $\mathbb{M}$ with a countable measure, where the scale $s = 2^j$ and time translation $\tau = 2^{-j}k$. If the sequence $\{\phi_{j,k}\}$ is a frame for the Hilbert space $\mathbb{H}$ with a frame representation operator $\mathbb{L}$, then there are positive real scalars A and B such that

$$A||f||^2_{\mathbb{H}} \le ||\mathbb{L}f||^2_{\mathbb{M}} \le B||f||^2_{\mathbb{H}} \quad \forall f \in \mathbb{H}, \qquad (1)$$

where $\mathbb{L}f = \{\langle f, \phi_{j,k}\rangle : \ j,k \in \mathbb{Z}\}$ and $\langle x, y\rangle$ are the inner products of x and y, both belong to $\mathbb{H}$; and $||\mathbb{L}f||_{\mathbb{M}} = \sqrt{\sum_j \sum_k |\langle f, \phi_{j,k}\rangle|^2}$ is a candidate norm.

The above relationship is a norm equivalence and represents the degree of coherence of the signal f with respect to the frame set of scaling functions. It may be interpreted as enforcing an approximate energy transfer between the domains $\mathbb{H}$ and $\mathbb{L}(\mathbb{H})$. In other words, for all signals $f \in \mathbb{H}$, a scaled amount of energy is

* Supported in part by the U.S. Army Research Laboratory and the U.S. Army Office under Grant No DAAD19-01-1-0646.
** To whom correspondence should be addressed. Email: axr2@psu.edu

distributed in the coefficient domain where the scale factor lies between A and B.[9] However, the energy distribution is dependent on the signal's degree of coherence with the underlying frame $\{\phi_{j,k}\}$. For a signal f, which is coherent with respect to the frame $\{\phi_{j,k}\}$, norm equivalence in the frame representation necessarily implies that a few coefficients contain most of the signal energy and hence have relatively large magnitudes. Similarly, pure noise signal w being incoherent with respect to the set $\{\phi_{j,k}\}$, must have a frame representation in which the noise energy is spread out over a large number of coefficients. Consequently, these coefficients have a relatively small magnitude.[10]

Let $\tilde{f}$ be a noise corrupted version of the original signal f expressed as

$$\tilde{f} = f + \sigma\ w, \tag{2}$$

where w is additive white gaussian noise with zero mean and unit variance and σ is the noise level. Then, the inner product of $\tilde{f}$ and $\phi_{j,k}$ is obtained as

$$\langle \tilde{f}, \phi_{j,k} \rangle = \underbrace{\langle f, \phi_{j,k} \rangle}_{\text{signal part}} + \sigma\ \underbrace{\langle w, \phi_{j,k} \rangle}_{\text{noise part}}. \tag{3}$$

The noise part in Eq. (3) may further be reduced by appropriate choice of the scales over which coefficients are obtained.

For every wavelet, there exists a certain frequency called the center frequency F_c that has the maximum modulus in the Fourier transform of the wavelet.[11] The pseudo-frequency f_p of the wavelet at a particular scale α is given by the following formula:[11,12]

$$f_p = \frac{F_c}{\alpha \Delta t}, \tag{4}$$

where Δt is the sampling interval. It is observed from numerical simulation that the signal to noise ratio can be improved by a factor of 6 or more through appropriate selection of the wavelet basis function and scales.

The power spectral density (PSD) of the signal provides the information about the frequency content of the signal. This information along with Eq. (4) can be used for scale selection. The procedure for selecting the scales is summarized as follows: (i) identification of the frequencies of interest through PSD analysis of time series data, (ii) substitution of the above frequencies in place of f_p in Eq. (4) to obtain the respective scales in terms of the known parameters F_c and Δt.

The wavelet coefficients of the signal are significantly large when the pseudo-frequency f_p of the wavelet corresponds to the locally dominant frequencies in the underlying signal. The wavelet coefficients are obtained upon selection of the wavelet basis and scale range. These coefficients are stacked at selected time-shift positions, starting with the smallest value of scale and ending with its largest value and then back from the largest value to the smallest value of the scale at the next instant of time shift. In the sequel, this one-dimensional array of arranged wavelet coefficients is called the *scale series* data, which is structurally similar to time series data in the phase space. For symbol generation, the scale series data are handled in a similar way as time series data.

The objective is to convert the scale series data to a sequence of symbols. A scheme for partitioning based on thresholds was used in Ref. [1]. In this approach, the maximum and minimum of the scale series are evaluated. The ordinates between the maximum and minimum are divided into equal-sized regions. These regions are obviously mutually disjoint and thus form a partition. Each region is then labelled with one symbol from the alphabet. If the data point lies in a particular region, it is coded with the symbol associated with that region. Thus, a sequence of symbols is created from a given scale series data set. This is called the uniform partitioning that is also reported in Refs.[1,8].

In an alternative form of partitioning, the regions with rich information are partitioned finer than those with sparse information. This objective is achieved by maximization of entropy that induces uniform probability distribution of the symbols in the alphabet Σ. Thus, small changes in the behaviour of the underlying dynamical system are more likely to be captured from the symbol sequence obtained under maximum entropy partitioning than under another partitioning. The procedure for maximum entropy partitioning that has been adopted in the proposed WS partitioning is succinctly described in the following.

Let N be the data set length and $|\Sigma|$ be the alphabet size, i.e., the number of (disjoint) members in the partition. The data set to be partitioned is sorted in ascending order. Starting from the first point in the sorted data set, every consecutive data segment of length $\lfloor \frac{N}{|\Sigma|} \rfloor$ forms a distinct member of the partition. Here $\lfloor x \rfloor$ represents the greatest integer less than or equal to x. Choice of the alphabet size $|\Sigma|$ also plays a vital role in information extraction, and the concept of entropy rate has been used for this purpose, as explained in the following.

Let $H(k)$ denote the entropy of the symbol sequence obtained by partitioning the data set with k symbols.

$$H(k) = -\sum_{i=1}^{i=k} p_i \log_2 p_i, \tag{5}$$

where p_i represents the probability of occurrence of the symbol σ_i and $H(1) = 0$. The maximum entropy of the data set is $\log_2(k)$ corresponding to the uniform probability distribution of the k symbols. The entropy rate, with respect to the number of symbols, is given by

$$h(k) = H(k) - H(k-1), \quad \forall k \geq 2. \tag{6}$$

An algorithm for choosing the alphabet size is given as follows: (1) Set $k = 2$. Choose a threshold ϵ_h. (2) Sort the set of scale series data set (of length N) in the ascending order. (3) Every consecutive segment of length $\lfloor \frac{N}{|\Sigma|} \rfloor$ in the sorted data set forms a distinct element of the partition. (4) Convert the scale series sequence to a symbol sequence with the partitions obtained in step (3). If the data point lies within or on the lower bound of a partition, it is coded with the symbol associated with that partition. (5) Compute the symbol probabilities p_i, $i = 1, 2, \cdots, k$. (6) Compute the entropy $H(k) = -\sum_{i=1}^{i=k} p_i \log_2 p_i$ and compute the entropy rate $h(k) = H(k) - H(k-1)$. (7) If $h(k) < \epsilon_h$, then exit; else increment k by 1 and go to step (3).

The choice of the threshold ϵ_h depends on the signal characteristics and may vary for individual systems. A small ϵ_h leads to a large size of the symbol alphabet, resulting in increase of computation. On the other hand, a large ϵ_h may fail to capture the small changes in dynamics because of excessively coarse graining resulting from a reduced number of symbols.

The performance of WS partitioning is now evaluated by comparison with SFNN partitioning in the context of symbolic time series analysis (STSA) for anomaly detection.[1] The objective of anomaly detection is to identify small changes in the critical parameter(s) as early as possible and well before it manifests as a drastic change in the behaviour of the dynamical system. Next, we present some application examples based on the time series data generated from two laboratory apparatuses.

The first example is built upon on experimental data, generated from a nonlinear electronic system apparatus[13] that emulates the forced Duffing equation:[14]

$$\frac{d^2y}{dt^2} + \beta\frac{dy}{dt} + y(t) + y^3(t) = A\cos(\Omega t), \tag{7}$$

where the dissipation parameter β varies slowly with respect to the response of the dynamical system; $\beta = 0.1$ represents the nominal condition; and a change in the value of β is considered as an anomaly. With amplitude $A = 22.0$ and $\Omega = 5.0$, a sharp change in the behaviour is noticed to be around $\beta = 0.29$, possibly due to bifurcation. The phase plots and time-response plots, depicting this drastic change behaviour, are not presented here because they are available in earlier publications.[1,2]

The same sets of time series data of the signal $y(t)$, generated from the nonlinear electronic system apparatus, have been used for symbolic analysis in SFNN and WS partitioning. In both the cases, the symbol alphabet size is chosen to be $|\Sigma| = 8$; the Gaussian wavelet, gaus1, is chosen as the basis function for WS partitioning. Once the partition is generated at the nominal condition of $\beta = 0.1$, it is kept invariant for subsequent analysis at different values of β. As the dynamical behaviour of the system changes due to variations in β, the statistical characteristics of the symbol sequences are also altered and so do the symbol probabilities. A measure is induced on deviations of the symbol probability vectors obtained under different anomalous conditions, to quantify these changes. Such a measure is called the anomaly measure $\boldsymbol{M}$. The metric $\boldsymbol{M}_k = d(\boldsymbol{p}_0, \boldsymbol{p}_k)$ is an anomaly measure, where $\boldsymbol{p}_0$ and $\boldsymbol{p}_k$ represent the symbol probability vectors under nominal and anomalous conditions, respectively. A candidate anomaly measure is the angle between the symbol probability vectors under nominal and anomalous conditions. This measure is defined as

$$\boldsymbol{M}_k = \arccos\left(\frac{\langle \boldsymbol{p}_0, \boldsymbol{p}_k \rangle}{||\boldsymbol{p}_0||_2 ||\boldsymbol{p}_k||_2}\right), \tag{8}$$

where $\langle x, y \rangle$ is the inner product of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$; and $||x||_2$ is the Euclidean norm of $\boldsymbol{x}$.

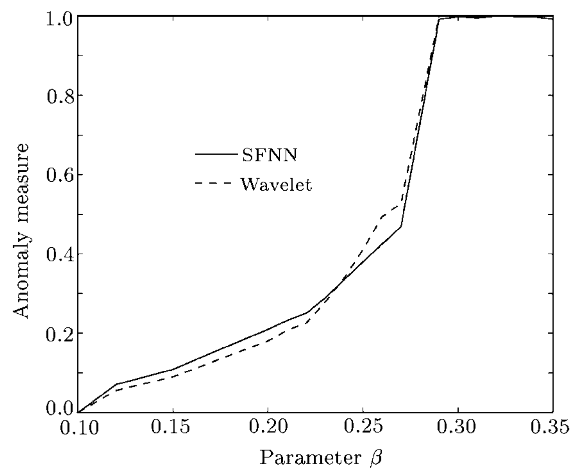

Fig. 1. Anomaly Detection on the electronic system apparatus.

Figure 1 depicts two plots of anomaly measure versus the dissipation parameter β, which are obtained from the same data sets by SFNN partitioning and WS partitioning. A comparison of the two plots in Fig. 1 reveals that the anomaly measure profile derived by WS partitioning is very close to that derived by SFNN partitioning. The execution time for SFNN to generate the partition is found to be about four hours, while that for WS partitioning is ~ 100 ms on the same computer with identical memory requirement. Therefore, it is logical to infer that, for this class of nonlinear electronic systems, WS partitioning is computationally several orders of magnitude more efficient than SFNN partitioning while both the partitioning methods yield similar performance from the perspective of anomaly detection.

The second example is built upon the experimental data, generated from a multi degree-of-freedom mechanical vibration system apparatus.[15] Figure 2 depicts the experimental apparatus.

The apparatus is persistently excited at a frequency of 10.4 Hz, which is a close approximation of one of resonance frequencies of the mechanical structure. Each set of time series data contains 30 s of information under persistent vibratory motion of the mass-beam system. The resulting cyclic stresses induce (irreversible) fatigue crack damage in the critical structures, which cause gradual reduction in stiffness. Consequently, the statistics of time series data undergo changes. The objective here is to detect these changes as early as possible in real time.

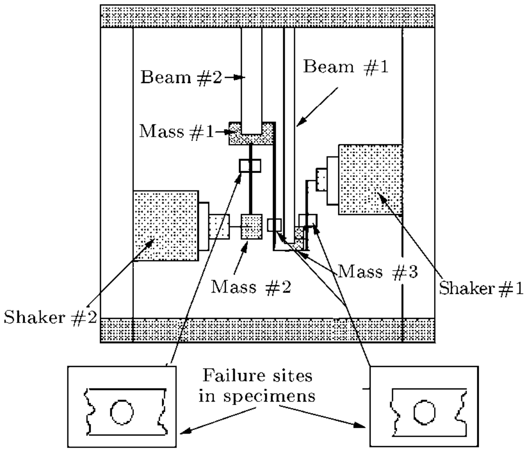

Fig. 2. Multi-degree-of-freedom mechanical vibration system apparatus.

The first data set, which is dominated by a sinusoid of frequency $\sim 10.4\,\mathrm{Hz}$ and represents the nominal behaviour of the mechanical vibration system, is considered to be the reference point.

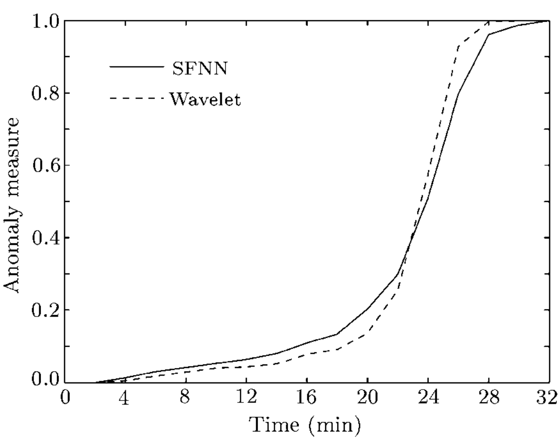

Fig. 3. Anomaly detection on the mechanical vibration system apparatus.

Figure 3 depicts the two plots of anomaly measure versus time, which are obtained by SFNN partitioning and WS partitioning from the same data sets and the same alphabet size $|\Sigma| = 8$. The gaussian wavelet, gaus9, is chosen as the basis function for WS partitioning. It is observed from the two anomaly measure profiles in Fig. 3 that WS partitioning is comparable to SFNN partitioning. Similar to the previous example in Fig. 1, the execution time for WS partitioning is about five orders of magnitude less than that for SFNN partitioning with identical memory requirement. Hence, it is inferred from this experiment that WS partitioning, is better suited for real-time detection of structural degradation in mechanical vibration systems. Another advantage of wavelet-based partitioning is that its effectiveness with noisy time series data. It is observed that the effects of noise can be significantly mitigated by appropriate selection of the wavelet basis function and scales.

A major conclusion based on this investigation is that wavelet-based (WS) partitioning, combined with appropriate choices of a wavelet basis function, significantly enhances computational efficiency and anomaly detection capabilities higher than those reported in literature. The field of symbolic time series analysis is relatively new and its application to anomaly detection is very recent. Therefore, the proposed method of partitioning for symbol generation requires continued theoretical and experimental research. Future research is recommended in the following areas: (1) exploration of lifting techniques[16] for wavelet customization, (2) extension of the proposed WS partitioning to multi-dimensional time series, (3) blind noise separation[5,6] in the time series data for robust partitioning.

References

[1] Ray A 2004 *Signal Processing* **84** 1115
[2] Chin S, Ray A and Rajagopalan V 2005 *Signal Processing* **85** 1859
[3] Gupta S, Ray A and Keller E 2006 *Mechanical Systems and Signal Processing* (Amsterdam: Elsevier)
[4] Daw C S, Finney C E A and Tracy E R 2003 *Rev. Sci. Instrum.* **74** 915
[5] Abarbanel H D I 1996 *The Analysis of Observed Chaotic Data* (New York: Springer)
[6] Kantz H and Schreiber T 2004 *Nonlinear Time Series Analysis* 2nd edn (Cambridge: Cambridge University Press)
[7] Kennel M B and Buhl M 2003 *Phys. Rev. Lett.* **91** 084102
[8] Gamero L G, Plastino A and Torres M E 1997 *Physica* **246** 487
[9] Mallat S 1998 *A Wavelet Tour of Signal Processing* 2nd edn (New York: Academic)
[10] Teolis A 1998 *Computational Signal Processing with Wavelets* (Berlin: Birkhäuser)
[11] *Wavelet Toolbox: Matlab* (Mathworks Inc.)
[12] Abry P 1997 *Ondelettes et Turbulence, Multirsolutions, Algorithmes de Dcomposition Invariance D'chelles* (Paris: Diderot Editeur)
[13] Rajagopalan V, Samsi R, Ray A, Mayer J and Lagoa C *A Symbolic Dynamics Approach for Early Detection of Slowly Evolving Faults in Nonlinear Systems* (IASTED-CSS, Clearwater, FL, November–December 2004)
[14] Thompson J M T and Stewart H B 1986 *Nonlinear Dynamics and Chaos* (Chichester: Wiley)
[15] Khathate, Ray A, Chin S, Rajagopalan V and Keller E 2004 *Proceedings of American Control Conference* (Boston, MA, June–July 2004) 3746
[16] Sweldens W 1998 *SIAM J. Math. Anal.* **29** 511