Space partitioning via Hilbert transform for symbolic time series analysis

Aparna Subbu[a)] and Asok Ray[b)]
The Pennsylvania State University, University Park, Pennsylvania 16802, USA

(Received 18 December 2007; accepted 28 January 2008; published online 28 February 2008)

Symbol sequence generation is a crucial step in symbolic time series analysis of dynamical systems, which requires phase-space partitioning. This letter presents analytic signal space partitioning (ASSP) that relies on Hilbert transform of the observed real-valued data sequence into the corresponding complex-valued analytic signal. ASSP yields comparable performance as other partitioning methods, such as symbolic false nearest neighbor partitioning (SFNNP) and wavelet-space partitioning (WSP). The execution time of ASSP is several orders of magnitude smaller than that of SFNNP. Compared to WSP, the ASSP algorithm is analytically more rigorous and is approximately five times faster. © *2008 American Institute of Physics*.
[DOI: 10.1063/1.2883958]

Symbolic time series analysis (STSA) has been proposed for real-time anomaly detection in complex systems.[1,2] A crucial step in STSA is partitioning of the phase space of the underlying dynamical system for symbol sequence generation.[3] Several techniques have been suggested in the physics literature for symbol generation, such as symbolic false nearest neighbors partitioning[4] (SFNNP) and wavelet-space partitioning (WSP).[1,5]

SFNNP optimizes a *generating partition* by avoiding topological degeneracy. The optimizing criterion in SFNNP is that short sequences of consecutive symbols should localize the corresponding state space point as closely as possible. This is achieved by forming a 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 degeneracy where symbolic words map back to globally distinct regions of the state space. The nearest neighbor to each point in the embedding is described in terms of the Euclidean distance between symbolic neighbors. Thus, better partitions yield a smaller proportion of symbolic false nearest neighbors. For convenience of implementation, the partitions are parametrized 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 neighbors is minimized over the free parameters using a genetic algorithm suitable for continuous-parameter spaces. A major shortcoming of SFNNP 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, the symbolic false neighbors rapidly grow in number and may erroneously require a large symbol alphabet to capture pertinent information on the system dynamics.

The wavelet transform largely alleviates the above shortcomings and is particulary effective with noisy data from high-dimensional dynamical systems. The WSP (1 and 5) was introduced as an alternative to SFNNP, where the basis and scales of the wavelet are determined from the power spectral density of the observed data. The wavelet coefficients at selected scale(s) are stacked back to back to transform the two-dimensional scale-shift wavelet domain into a one-dimensional domain. The resulting scale-series data sequence is converted to a sequence of symbols by maximum entropy partitioning.[5] Although WSP is significantly computationally faster than SFNNP and is suitable for real-time applications, WSP has several shortcomings that include the following.

- The selection of an appropriate wavelet basis function is made based on inspection of the power spectral density of the underlying signal, which may vary with the window size. Apparently, there is no precise way of selecting a wavelet basis that is "best" for partitioning.
- The identification of scales for generation of wavelet coefficients are identified from the center frequency (that is also based on visual inspection of the power spectral density of the Fourier transform) and the selected wavelet basis.
- The dimension reduction of the scale-shift wavelet domain to a one-dimensional domain of scale-series sequences is nonunique and may not be a "best" way.

This letter presents a partitioning method, called analytic signal space partitioning (ASSP), for STSA as an alternative to the existing partitioning methods. The purpose of ASSP is to capture the relevant statistical patterns for anomaly detection in real time. Although ASSP is not aimed to be a generating partition, it is designed with the goal of satisfying the important property of a generating partition: the inverse image of a small neighborhood in the symbol space is a small neighborhood in the data space, except possibly in the vicinity of partition boundaries. The contributions of the work reported in this letter are concept development, formulation, and validation of ASSP with experimental data from a laboratory apparatus.

The underlying concept of ASSP partitioning is built upon Hilbert transform of the observed real-valued data sequence into the corresponding complex-valued analytic signal[6] as explained below.

[a)]Electronic mail: avs127@psu.edu.
[b)]Electronic mail: axr2@psu.edu.

Let $x(t)$ be a real-valued function whose domain is the real field $\mathbb{R}=(-\infty,+\infty)$. Then, Hilbert transform[7] of $x(t)$ is defined as

$$\tilde{x}(t) = \mathcal{H}[x](t) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{x(\tau)}{t-\tau} d\tau. \tag{1}$$

That is, $\tilde{x}(t)$ is the convolution of $x(t)$ with $1/\pi t$ over $\mathbb{R}$, which is represented in the Fourier domain as

$$\mathcal{F}[\tilde{x}](\xi) = -i\,\mathrm{sgn}(\xi)\mathcal{F}[x](\xi),$$

$$\text{where}\ \ \mathrm{sgn}(\xi) = \begin{cases} +1 & \text{if } \xi > 0 \\ -1 & \text{if } \xi < 0 \end{cases}. \tag{2}$$

Given the Hilbert transform of a real-valued signal $x(t)$, the corresponding complex-valued analytic signal is defined as

$$\mathcal{A}[x](t) = x(t) + i\tilde{x}(t). \tag{3}$$

The construction of Eq. (3) is based on the fact that the values of Fourier transform of a real-valued function at negative frequencies are redundant due to their Hermitian symmetry imposed by the transform. Thus, the phase of the Hilbert transform $\tilde{x}(t)$ is in quadrature to the phase of $x(t)$. That is, the analytic signal can be expressed as

$$\mathcal{A}[x](t) = A(t)\exp[i\varphi(t)], \tag{4}$$

where $A(t)$ and $\varphi(t)$ are called the instantaneous amplitude and instantaneous phase of $\mathcal{A}[x](t)$, respectively. Vakman[8] has pointed out that the amplitude and phase of an analytic signal satisfy the following three physical properties.

(1) Amplitude continuity is a small perturbation in $x(t)$ inducing a small change in $A(t)$.
(2) Phase independence of scale is scaling $x(t)$ by a constant $c>0$, which has no effects on $\varphi(t)$ and multiplies $A(t)$ by c.
(3) Harmonic correspondence is a monofrequency signal (i.e., a pure sinusoid $A_0\cos[\omega_0 t+\varphi_0]$) yielding $A(t)=A_0$ and $\varphi(t)=\omega_0 t+\varphi_0$ for all t.

Thus, for a monofrequency signal, which is embedded in a two-dimensional state space, a direct parallel can be drawn between the phase plot and the Hilbert transform plot. The procedure for ASSP is formulated next.

Let the observed signal be available as a real-valued time series of N data points. Upon Hilbert transformation of this data sequence, a pseudophase plot is constructed from the resulting analytic signal by a bijective mapping of the complex field onto $\mathbb{R}^2$, i.e., by plotting the real and the imaginary parts of the analytic signal on the x_1 and x_2 axes, respectively. It is important to note that the pseudophase space is always two-dimensional, whereas the phase space of the dynamical system is a representation of the n-dimensional manifold, where n could be an arbitrarily large positive integer.

The time-dependent analytic signal in Eq. (3) is now represented as a (one-dimensional) trajectory in the the two-dimensional pseudophase space. Let Ξ be a compact region in the pseudophase space, which encloses the trajectory. The objective here is to partition Ξ into finitely many mutually exclusive and exhaustive segments, where each segment is labeled with a symbol or letter. The segments are conveniently determined by the magnitude and phase of the analytic signal as well as based on the density of data points in these segments. That is, if the magnitude and phase of a data point of the analytic signal lies within a segment or on its boundary, then the data point is labeled with the corresponding symbol. Thus, a symbol sequence is naturally derived from the (complex-valued) sequence of the analytic signal. The set of (finitely many) symbols is called the alphabet Σ.

One possible way of partitioning Ξ would be to divide the magnitude and phase of the time-dependent analytic signal in Eq. (3) into uniformly spaced segments between their maximum and minimum values, respectively. This is called uniform partitioning. An alternative method, known as maximum entropy partitioning,[5] maximizes the entropy of the partition that is characterized by the alphabet size $|\Sigma|$, thereby imposing a uniform probability distribution on the symbols. In this partitioning, parts of the state space with rich information are partitioned into finer segments than those with sparse information. Computationally, the maximum entropy partition can be obtained by sorting the data sequence in an ascending order. This sorted data sequence is then partitioned into $|\Sigma|$ equal segments of length $[N/|\Sigma|]$, where N is the length of the data sequence and $[x]$ is the greatest integer less than or equal to x. Each of these segments is assigned a symbol and all data points in a given segment are assigned the corresponding symbol.

The magnitude and phase of the analytic signal in Eq. (3) are partitioned separately according to either uniform partitioning, maximum entropy partitioning, or any other type of partitioning; the type of partitioning may depend on the characteristics of the physical process. In essence, each point in the data set is represented by a pair of symbols—one belonging to the alphabet Σ_R based on the magnitude (i.e., in the radial direction) and the other belonging to the alphabet Σ_A based on the phase (i.e., in the angular direction). The analytic signal is partitioned into a symbol sequence by associating each pair of symbols into a symbol from a new alphabet Σ as

$$\Sigma \triangleq \{(\sigma_i,\sigma_j):\sigma_i \in \Sigma_R, \sigma_j \in \Sigma_A\}\ \text{and}\ |\Sigma| = |\Sigma_R||\Sigma_A|.$$

The construction of the ASSP is now complete. Next, the performance of ASSP is evaluated against SFNNP and WSP in the context of STSA for anomaly detection,[1] where the objective is to identify small changes in the critical parameters of a dynamical system as early as possible before it manifests into a catastrophic disruption (e.g., onset of a chaos in the dynamical system sense or a phase transition in the thermodynamic sense) in the behavior of the dynamical system.

We present an example based on the time series data, generated from a laboratory apparatus of nonlinear active electronic systems, which emulates the forced Duffing equation

$$\frac{d^2y}{dt^2} + \beta(t_s)\frac{dy}{dt} + y(t) + y^3(t) = A\cos(\Omega t), \tag{5}$$

where the dissipation parameter β varies in the slow time scale t_s with respect to the time scale t of the dynamical system; $\beta=0.1$ represents the nominal condition. A change in the value of β from its nominal value is considered as an anomaly. For $A=22.0$ and $\Omega=5.0$, a sharp change in

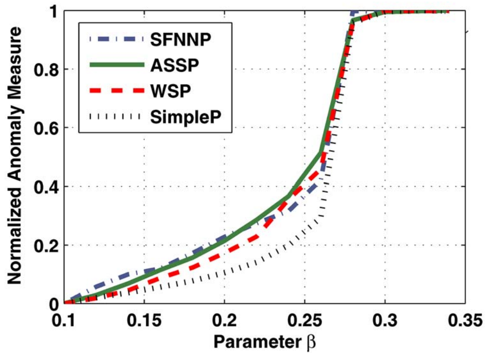

FIG. 1. (Color online) Comparison of partitioning methods for anomaly detection.

the behavior occurs around $\beta \approx 0.29$ possibly due to a bifurcation.[1,2,5]

The capability for anomaly detection in the Duffing system was evaluated for different quasistatic values of the slowly varying parameter β. For each of the time series data sets, corresponding to a value of β, symbolic analysis for anomaly detection was performed using SFNNP, WSP, and ASSP. This letter also shows the results by maximum entropy partitioning of time series data without any transformation and embedding, referred to as simple partitioning (simple*P*); that is, simple*P* is executed only on the real part of the complex data on which ASSP is executed. The objective here is to evaluate the performance of ASSP relative to SFNNP, WSP, and simple*P* for detection of small anomalies.

The symbol alphabet size $|\Sigma|=8$ was chosen for each of SFNNP and WSP and $|\Sigma|=5$ for simple*P*. For ASSP, alphabet sizes were chosen as $|\Sigma_R|=5$ in the radial direction and $|\Sigma_A|=3$ in the angular direction. For each of the four cases, the partition was constructed based on the same set of observed data and the respective partitions were kept invariant for all data sets at subsequent values of β. A deviation in β affects the dynamical behavior of the nonlinear system and thereby the underlying statistics of the derived symbol sequences change. In this context, a measure of performance of the partitioning methods for anomaly detection was constructed based on the probability distribution $\mathbf{p}$ of symbol occurrence, which is treated as a pattern vector.

One possible measure of anomaly, which is adopted in this letter, is the angle between pattern vectors at the nominal and anomalous conditions. The anomaly measure at the kth epoch is defined as

$$\mathbf{M}_k = \arccos\left(\frac{\langle p_0, p_k\rangle}{\|p_0\|_2\|p_k\|_2}\right), \tag{6}$$

where $\langle p_0, p_k\rangle$ is the inner product of probability vectors p_0 and p_k at the nominal condition and the kth epoch, respectively; and $\|\cdot\|_2$ is the Euclidian norm of $\cdot$. The plots in Fig. 1 depict the anomaly measure profile for each of the four partitioning methods: SFNNP, WSP, ASSP, and simple*P*. While the results are qualitatively similar for all four cases, simple*P* is outperformed by the other partitioning methods including ASSP that is executed on a more rich source of information than simple*P*. The execution time for SFNNP was over 4 h; in contrast, the execution time was ~ 1.5 and ~ 0.3 s for WSP and ASSP, respectively, and ~ 0.25 s for simple*P* that is slightly faster than ASSP.

To validate the above experimental results for higher order dynamical systems, a simulation test bed has been constructed with four van der Pol oscillators in cascade. The first oscillator provides a sinusoidal input to the second oscillator and the output of the last oscillator in the cascade is fed back to the second oscillator to simulate a sixth order nonautonomous dynamical system. Simulation experiments were conducted on the test bed for performance comparison of WSP, ASSP, and simple*P*, where simple*P* was significantly outperformed by ASSP and WSP. Because of excessive computation time requirements, SFNNP was not investigated on the simulation test bed. The performances of WSP and ASSP were quite similar and the ratio of their respective execution time was also similar (i.e., ~ 5) to what was observed earlier for the Duffing system.

It is concluded that the usage of Hilbert transform and analytic signals provides a superior partitioning method in symbolic time series analysis for real-time anomaly detection.[1] The results generated by the proposed ASSP method, as applied to electronic circuits on a laboratory apparatus for anomaly detection, are similar to those generated by SFNNP (4) with several orders of magnitude smaller computational cost of ASSP. As such, SFNNP serves the role of a benchmark for testing, evaluation, and calibration of other partitioning algorithms. Due to its natural Fourier domain interpretation, ASSP is more easily implementable and is physically more intuitive than the WSP (5) that requires a good understanding of the signal characteristics for selection of the wavelet basis, identification of appropriate scales, and conversion of the two-dimensional scale-shift domain into a single dimension. However, for noise-corrupted data, WSP has the advantage of having the inherent capability of denoising.

Future research areas for enhancement of the ASSP algorithm include:

- development of a rigorous algorithm to determine the number of symbols in both radial and angular directions; and
- investigation of effectiveness of the partitioning scheme for non-Markov (e.g., with long memory) systems.

This work has been supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under Grant No. W911NF-07-1-0376 and by NASA under Cooperative Agreement No. NNX07AK49A.

[1] A. Ray, Signal Process. **84**, 1115 (2004).
[2] S. Gupta and A. Ray, Appl. Phys. Lett. **91**, 94105 (2007).
[3] C. Daw, C. Finney, and E. Tracy, Rev. Sci. Instrum. **74**, 915 (2003).
[4] M. Buhl and M. Kennel, Phys. Rev. E **71**, 046213 (2005).
[5] V. Rajagopalan and A. Ray, Signal Process. **86**, 3309 (2006).
[6] D. Gabor, J. Inst. Electr. Eng., Part 1 **93**, 429 (1946).
[7] L. Cohen, *Time-Frequency Analysis* (Prentice-Hall, Upper Saddle River, NJ, 1995), Chap. 2, pp. 30–31.
[8] L. Cohen, P. Loughlin, and D. Vakman, Signal Process. **79**, 301 (1999).