

Pattern identification in dynamical systems via symbolic time series analysis[☆]

Venkatesh Rajagopalan, Asok Ray*, Rohan Samsi, Jeffrey Mayer

The Pennsylvania State University, University Park, PA 1802, USA

Received 16 January 2006; received in revised form 28 February 2007; accepted 5 March 2007

Abstract

This paper presents symbolic time series analysis (*STSA*) of multi-dimensional measurement data for pattern identification in dynamical systems. The proposed methodology is built upon concepts derived from *Information Theory* and *Automata Theory*. The objective is not merely to classify the time series patterns but also to identify the variations therein. To achieve this goal, a symbol alphabet is constructed from raw data through partitioning of the data space. The maximum entropy method of partitioning is extended to multi-dimensional space. The resulting symbol sequences, generated from time series data, are used to model the dynamical information as finite state automata and the patterns are represented by the stationary state probability distributions. A novel procedure for determining the structure of the finite state automata, based on entropy rate, is introduced. The diversity among the observed patterns is quantified by a suitable measure. The efficacy of the *STSA* technique for pattern identification is demonstrated via laboratory experimentation on nonlinear systems.

© 2007 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

Keywords: Pattern classification; Symbolic time series analysis; Markov modeling

1. Introduction

Accurate modeling of system dynamics, based solely on the fundamental principles of physics, is often infeasible. Therefore, it might be necessary to rely on time series data generated from sensors to learn about the system dynamics. When the dynamics are stationary, analytical methods like the Fourier transform are adequate for recognizing the patterns in the time series. However, if the signal is nonstationary (e.g., exhibiting drift and/or frequency variations), the afore-mentioned analytical tools may become inadequate. Moreover, nonlinear and nonstationary systems often exhibit behavior like strange attraction, chaos, and bifurcation [1]. In such cases, a more powerful technique is necessary for pattern classification and also for identifying the variations therein [2].

Symbolic time series analysis (*STSA*) is a useful tool for modeling and characterization of nonlinear dynamical systems [3,4].

A novel approach, based on symbolic dynamics, for anomaly detection was introduced by Ray in Ref. [5]. This methodology was compared with other data driven methods such as principal component analysis (*PCA*) and artificial neural networks (*ANN*) in Ref. [6]. A new partitioning scheme to enhance this methodology, known as maximum entropy (ME) partitioning, was reported in Ref. [7]. The *STSA* methodology is briefly summarized below.

Time series data are converted to symbol sequences based on an appropriate partition. This partition is obtained with respect to a time series chosen as the nominal. The partition remains invariant in the analysis of subsequent data sets. As the dynamical behavior of the system changes, symbol sequences generated are expected to be different from those of the nominal. Probabilistic finite state automata (*PFSA*) [8] can be used to model the dynamics of the symbolic process generated from the time series. The probability distributions, obtained from the *PFSA*, provide a statistical representation of the patterns. The variations in symbolic patterns are quantified by the divergence among the probability distributions.

If the partitioning of time series and the subsequent modeling are done appropriately, the symbolic dynamic analysis can

[☆] This work has been supported in part by the US Army Research Laboratory and the US Army Research Office under Grant No. DAAD19-01-1-0646.

* Corresponding author. Tel.: +1 814 865 6377.

E-mail address: axr2@psu.edu (A. Ray).

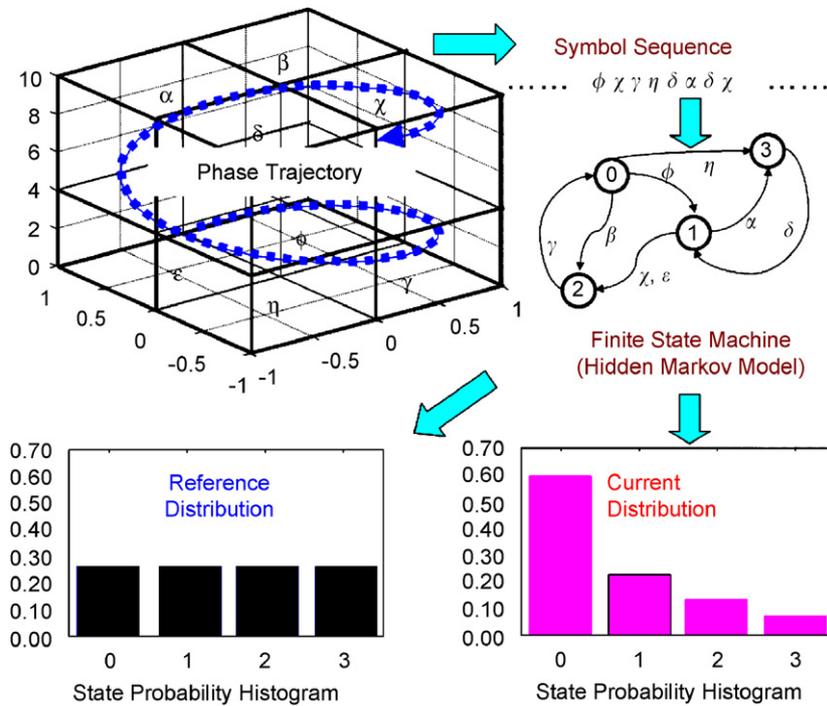


Fig. 1. Pattern recognition via symbolic time series analysis.

accurately identify and represent the patterns observed in the time series. Fig. 1 provides a pictorial depiction of the pattern identification process beginning with partitioning of the data space to generate symbols, followed by representation with *PFSA* and the generation of probability distributions from the *PFSA*. Moreover, there are several advantages associated with *STSA*. An important practical advantage of symbolic analysis is its enhanced computational efficiency. Analysis of symbolic data is often less sensitive to measurement noise. Applications of symbolic methods are thus favored in circumstances where robustness to noise and computational speed are paramount [4].

There are different classes of pattern identification (e.g., syntactic or structural matching, statistical decision-theoretic, and artificial neural networks) [9], which are based on raw data. The above classes may not be mutually exclusive as it is possible that the same pattern recognition method may belong to more than one class. The *STSA*-based pattern identification tool, presented in this paper, can be interpreted to belong to both syntactic and statistical decision-theoretic classes.

The paper is organized into six sections including the present one. Section 2 provides a brief introduction to *STSA*. Section 3 describes the *ME* partitioning [7] method for symbol generation. This method is extended to multiple dimensions in this paper. Pattern representation with finite state machines and measures for quantifying deviations are described in Section 4. Section 5 discusses the results of the proposed *STSA* methodology, as applied to the well known nonlinear systems described by the Duffing equation [10] and the Van der Pol equation [11]. Section 6 summarizes and concludes the paper with recommendations for future research.

2. Symbolic time series analysis

Continuously varying physical processes are often modeled as a finite-dimensional dynamical system:

$$\frac{dx(t)}{dt} = f(x(t), \theta); \quad x(0) = x_0, \quad (1)$$

where $t \in [0, \infty)$ is time; $x \in \mathbb{R}^\ell$ is the state vector in the phase space; and $\theta \in \mathbb{R}^m$ is the (slowly varying) parameter vector. Formally, a solution to Eq. (1) can be expressed as a continuous function of the initial state x_0 as $x(t) = \Phi_t(x_0)$, where Φ_t represents a parametric family of maps of the phase space into itself. This evolution of phase trajectory in (discrete) time t can be viewed as a flow of points in the phase space. A symbolic description is derived by partitioning the phase space into mutually disjoint regions as illustrated in Fig. 1. A brief discussion on partitioning follows.

A compact set $\Omega \in \mathbb{R}^\ell$, within which the trajectory is contained, is identified with the phase space itself. The encoding of Ω is accomplished by introducing a partition $\mathbb{B} \equiv \{B_0, B_1, B_2, \dots, B_{p-1}\}$ consisting of p mutually exclusive and exhaustive subsets, i.e., $B_j \cap B_k = \emptyset, \forall j \neq k, \bigcup_{j=1}^p B_j = \Omega$. Each phase trajectory is described by an orbit $O \equiv (x_0, x_1, x_2, \dots, x_\ell, \dots)$, which passes through or touches various elements of the partition \mathbb{B} . Let the index of domain $B_i \in \mathbb{B}$ visited at a given time instant be denoted by the symbol $\sigma_i \in \Sigma$. The set of symbols $\Sigma = (\sigma_0, \sigma_1, \dots, \sigma_{p-1})$ labeling the partition elements is called the alphabet. Each initial state x_0 generates an (infinite) sequence of symbols defined by a mapping from the phase space to the space

of symbols

$$x_0 \mapsto \sigma_3 \sigma_4 \sigma_2 \sigma_0 \dots \quad (2)$$

Such a mapping is called *symbolic dynamics* [3] as it attributes a legal (i.e., physically admissible) symbol sequence to the time series. Since the size of each cell is finite and also the cardinality of the alphabet Σ is finite, any such symbol sequence represents, through iteration, a phase trajectory that has the compact support Ω . In general, a dynamical system would only generate a subset of all possible sequences of symbols as there could be some illegal (i.e., physically inadmissible) sequences.

3. Symbol generation

The first aspect of *STSA* approach to pattern recognition is generation of symbol sequences from time series. Various methods have been suggested in literature for symbolization. These include variance-based [12] and entropy-based [13] methods as well as hierarchical clustering. A survey of various clustering techniques is provided in Ref. [14]. In addition to these methods, another scheme of partitioning, based on symbolic false nearest neighbors (*SFNN*), was introduced in Ref. [15]. The objective of *SFNN* partitioning is to ensure that points that are close to each other in symbol space are also close to each other in phase space. Partitions that yield a smaller proportion of *SFNN* are considered optimal. However, this partitioning method may become computationally very inefficient if the dimension of the phase space is large or if the data set is contaminated by noise, since noise induces false symbols. Rajagopalan and Ray [7] have reported comparison of computation speed of *SFNN* partitioning and wavelet-transform-based *ME* partitioning for measurement data, collected from three different laboratory apparatuses, namely, electronic system, mechanical vibration system, and fatigue damage system. In all three test apparatuses, the execution time of *ME* partitioning was found to be about five orders of magnitude less than that of *SFNN* partitioning on the same computer and for the same sets of data, while their performance in terms of anomaly detection matched very closely.

Entropy-based partitioning, introduced in Ref. [7] for one-dimensional data, is extended to multiple dimensions in this paper. A partition that maximizes the entropy of the generated symbol sequence is chosen as the candidate partition. In other words, this partition induces a uniform distribution of symbols for the nominal pattern. This method of *ME* partitioning is abbreviated as *ME* partitioning in the sequel. The procedure for obtaining an *ME* partition, for one-dimensional data, is described below.

3.1. ME partitioning for one-dimensional data

Let N be the length of the data set and $|\Sigma|$ be the size of the alphabet (i.e., the number of the disjoint elements in the partition). The data is sorted in ascending order. Starting from the first point in the sorted data, every consecutive data segment of length $\lfloor N/|\Sigma| \rfloor$ forms a distinct element of the partition. (Note: $\lfloor x \rfloor$ represents the greatest integer less than or equal to x .)

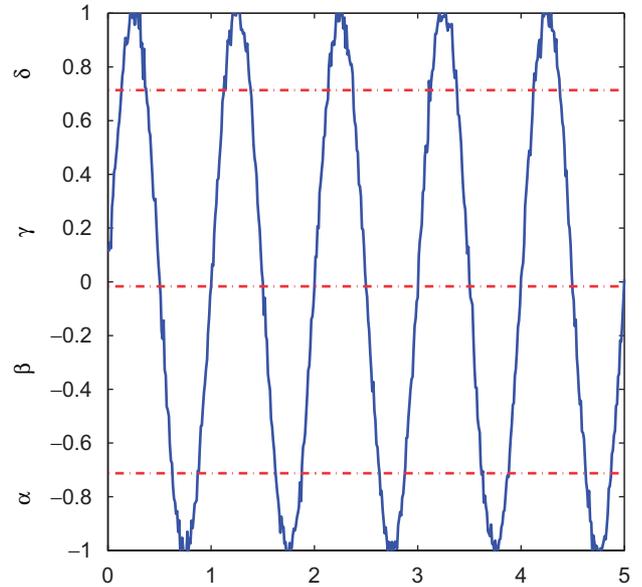


Fig. 2. *ME* partitioning with $|\Sigma| = 4$.

With *ME* partitioning, information-rich regions are allocated more symbols and hence a finer partition is achieved in such regions. Similarly, regions with sparse information content are allocated fewer symbols leading to a coarser partition in those regions. Hence, even small variations in patterns are more likely to be reflected in the symbol sequence obtained under *ME* partitioning than other partitioning. Fig. 2 shows an *ME* partitioning for the noise contaminated signal $\sin(2\pi t)$ with $|\Sigma| = 4$. As expected, the size of the partitions are not equal, but the probabilities of the symbols are equal.

The choice of the alphabet size $|\Sigma|$ plays a crucial role in *STSA*. For example, a small value of $|\Sigma|$ may prove inadequate for capturing the characteristics of the raw data. On the other hand, a large value may lead to redundancy and waste of computational resources. The selection of optimal Σ is an area of active research.

An entropy-based approach has been adopted for selecting the alphabet size. Let $H(k)$ denote the Shannon entropy of the symbol sequence obtained by partitioning the data set with k symbols.

$$H(k) = - \sum_{i=1}^{i=k} p_i \log_2 p_i, \quad (3)$$

where p_i represents the probability of occurrence of the symbol σ_i . Note that $H(1) = 0$ because $p_i = 0$ or 1 with $i = 1$. If the underlying data contain sufficient information content, then the entropy achieved under *ME* partitioning would be $\log_2(k)$, which corresponds to the uniform distribution. We define a quantity $h(\cdot)$ to represent the change in entropy with respect to the number of symbols.

$$h(k) \triangleq H(k) - H(k - 1) \quad \forall k \geq 2. \quad (4)$$

The algorithm for alphabet size selection is given below:

- Step 1: Set $k = 2$. Choose a threshold ϵ_h , where $0 < \epsilon_h \ll 1$.
- Step 2: Sort the data set (of length N) in the ascending order.

Step 3: Every consecutive segment of length $\lfloor N/k \rfloor$ in the sorted data set (of length N) forms a distinct element of the partition.

Step 4: Convert the raw data into 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.

Step 5: Compute the symbol probabilities $p_i, i = 1, 2, \dots, k$.

Step 6: Compute $H(k) = -\sum_{i=1}^k p_i \log_2 p_i$ and $h(k) = H(k) - H(k-1)$.

Step 7: If $h(k) < \varepsilon_h$, then exit; else increment k by 1 and go to Step 3.

A small value of threshold ε_h leads to a large size of the symbol alphabet, resulting in increased computation. Also a larger alphabet will make the partitioning finer. This might increase the probability of false symbols being induced by noise. On the other hand, a large ε_h will lead to a small alphabet size that may prove inadequate for pattern identification. Hence there is trade-off between accuracy and computational speed when ε_h is chosen. The variance of the noise process associated with the raw time series data may serve as a guideline for selection of ε_h .

3.2. ME partitioning for multi-dimensional data

While it is fairly simple to find a partition that satisfies the ME criterion for one-dimensional data, it is not straightforward to construct such a partition for multi-dimensional data. Some researchers have considered an approach where each dimension is partitioned independently. The cartesian products of these independent partitions form the elements of the partition for multi-dimensional space [13].

For example, in a p -dimensional space, the elements of a partition would assume the shape of a p -dimensional hypercube. This approach suffers from the limitation that every dimension should have the same number of alphabets. Such a restriction may not be appropriate as more information may be contained in a particular dimension than others and it may require a greater number of partitions.

Alternatively, different number of partitions may be chosen for various dimensions. But this would make the process tedious as the appropriate number of partitions needs to be determined for each dimension. Also the symbols obtained under such a partition may not be uniformly distributed. Example of such a case can be found in Ref. [13]. To overcome the above-mentioned limitations, the following approach is proposed.

Let $X = \{x_1, x_2, \dots, x_N\}$ be a data sequence of length N in a p -dimensional normed space. Define a functional $f(x)$ with a small number of free parameters. The value of these free parameters are determined from the data sequence $\{X\}$. The functional is evaluated for all data points as

$$s_i = f(x_i) \quad \forall x_i \in X. \quad (5)$$

The ME criterion is imposed on the sequence $\{s\}$ to obtain the partitions.

As an illustrative example, consider a two-dimensional data set. Let the functional chosen be

$$f(x) = \|x - \mu\|_2, \quad (6)$$

where $\|\cdot\|_2$ is the Euclidean norm in the two-dimensional vector space; μ is the two-dimensional centroid vector of the data defined as

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i. \quad (7)$$

The functional is evaluated for each data point x_i as

$$s_i = \|x_i - \mu\|_2. \quad (8)$$

The ME criterion is applied on the sequence $\{s\}$ to obtain the boundaries of each element of the partition. In a two-dimensional space these partitions appear as concentric circles with increasing radii. In higher dimensions, these would be hyper-spheroids. A data point x_i is coded with symbol $\sigma_j \in \Sigma$ if the condition $r_{j-1} \leq f(x_i) < r_j$ is satisfied, where r_{j-1} and r_j are the radii of two adjacent concentric circles.

It is essential that these partitions are defined appropriately taking into consideration, the dynamical information of the data being analyzed. It is highly unlikely that one type of partition will be suitable for all data. Further details are given later in Section 5.

4. Pattern representation with markov machines

PFSA are widely used in a variety of areas in pattern recognition and related fields (e.g., computational linguistics, machine learning, time series analysis, speech recognition, fault detection and machine translation). A survey of various properties of PFSA is provided in Ref. [8]. This section describes representation of symbol sequences with a special class of PFSA called the D -Markov machine [5]. It also presents appropriate measures for quantifying the diversity or similarity between a pair of patterns.

4.1. D -Markov machine construction

The core assumption in D -Markov machine construction is that a symbolic process can be approximated, to a desired level of accuracy, as a D th order Markov process. $D \in \mathbb{N}$ and \mathbb{N} is the set of natural numbers.

Definition 1. A stochastic symbolic process \mathbb{S} is called D th order Markov process if the probability of the current observation depends only on the previous D observations, i.e., $\forall k, \forall s_i \in \Sigma$

$$P[s_k | s_{k-1} s_{k-2} \dots s_1 s_0] = P[s_k | s_{k-1} s_{k-2} \dots s_{k-D+1} s_{k-D}].$$

In other words, the process has a memory of length D . Such a process can be represented as a PFSA. The states of the automaton are represented by symbol strings of length D , defined over the alphabet Σ . For example, with an alphabet $\Sigma = \{0, 1\}$

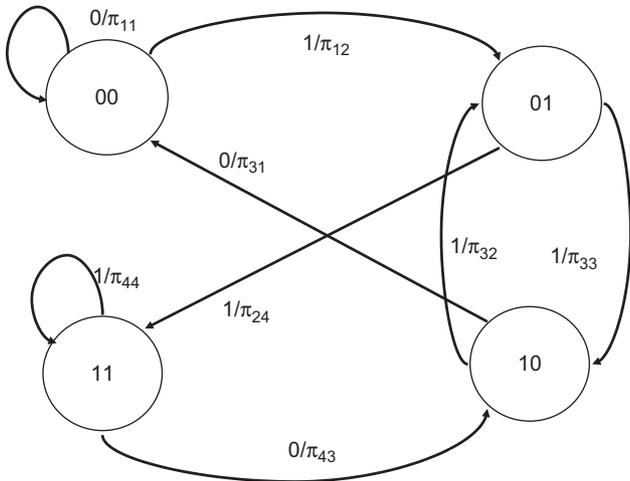


Fig. 3. D -Markov machine with $|\Sigma| = 2$ and $D = 2$.

and depth $D = 2$, the possible states are $\{00, 01, 10, 11\}$. The machine transitions from one state to another upon occurrence of a symbol $s_i \in \Sigma$. Fig. 3 depicts a D -Markov machine with $\Sigma = \{0, 1\}$ and $D = 2$.

The D -Markov machine can be represented by its state transition matrix (Π) or by its morph matrix $\tilde{\Pi}$. The Π -matrix is a stochastic matrix, i.e., all row sums are equal to one. The elements of the Π -matrix represent the probabilities of transition from one state to another. For example, π_{23} represents the probability of transition from state 2 to 3. The state transition probabilities of a D -Markov can be experimentally determined from the symbol sequence by frequency counting. The morph matrix $\tilde{\Pi}$, provides the conditional symbol probabilities for each state. For example, $\tilde{\pi}_{41}$ represents the probability of encountering symbol 1 when the machine is at state 4. While the Π -matrix is a square matrix, the $\tilde{\Pi}$ -matrix is a $|Q| \times |\Sigma|$ matrix. $|Q|$ is the number of states in the D -Markov machine. The Π -matrix and the $\tilde{\Pi}$ -matrix for the automaton represented in Fig. 3 are presented below:

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & 0 & 0 \\ 0 & 0 & \pi_{23} & \pi_{24} \\ \pi_{31} & \pi_{32} & 0 & 0 \\ 0 & 0 & \pi_{43} & \pi_{44} \end{bmatrix},$$

$$\tilde{\Pi} = \begin{bmatrix} \tilde{\pi}_{11} & \tilde{\pi}_{12} \\ \tilde{\pi}_{21} & \tilde{\pi}_{22} \\ \tilde{\pi}_{31} & \tilde{\pi}_{32} \\ \tilde{\pi}_{41} & \tilde{\pi}_{42} \end{bmatrix}.$$

The total number of states in a D -Markov machine is less than or equal to $|\Sigma|^D$ since some of the states might be forbidden, implying that the probabilities of these states are zero. Given the alphabet size $|\Sigma|$ and the depth D , states of the D -Markov machine are determined from the symbol sequence.

The depth of the D -Markov machine is a crucial parameter since the number of states varies exponentially with D . A very small depth could mean insufficient memory for the D -Markov

Table 1
Number of states and entropy rate for ideal string

Depth (D)	No. of states ($\leq \Sigma ^D$)	Entropy rate (h_μ)
0	1	0.810
1	2	0.689
2	3	0.500
3	4	0.000
4	4	0.000
5	4	0.000

machine to appropriately represent the symbolic dynamics of the process. On the other hand, an unnecessarily large D would result in a large number of states, leading to extremely small values of state probabilities and an inaccurate Π -matrix. A procedure based on entropy rate has been developed for selecting the depth of the D -Markov machine. The key idea is that increasing the depth beyond a certain value does not lead to any appreciable change in entropy; equivalently, the entropy rate would be very small.

Definition 2. Given the current state, the entropy rate, h_μ , of a symbolic stochastic process is defined as the uncertainty in the next symbol.

$$h_\mu = - \sum_{i=1}^N p_i \sum_{j=1}^{|\Sigma|} \tilde{\pi}_{ij} \log_2 \tilde{\pi}_{ij}, \tag{9}$$

where p_i is the probability of occurrence of i th state, $\tilde{\pi}_{ij}$ is the probability of occurrence of j th symbol in the i th state; N is the number of states in the probabilistic finite state machine; and $|\Sigma|$ is the alphabet size.

Being a measure of uncertainty, the entropy rate h_μ monotonically decreases as the depth D of the D -Markov machine is increased. Beyond a certain point, increasing D may not lead to any appreciable change in the entropy rate. This is the asymptotical entropy rate and the corresponding D is optimal for the machine. With ideal noise-free data h_μ converges to zero. However, in the real world of noisy data, h_μ may only monotonically decrease to a small nonzero value, depending on the magnitude and the type of noise. Thus, the test for the optimum D relies on how h_μ converges as D is increased. For example, let us consider a data set that yields a symbol stream $\bar{S} = \dots 000100010001 \dots$ on the alphabet $\Sigma = \{0, 1\}$. Table 1 provides the number of states and the entropy rate of the inferred D -Markov machine for various depths.

It can be seen that the number of states in the generated machine remains the same for depth $D \geq 3$. Correspondingly, the entropy rate remains at zero. This implies the minimum depth for correct representation for this symbol stream is 3. The number of states is less than $|\Sigma|^D$ in this case. The curve, shown in Fig. 4 by solid lines, exhibits the plot of h_μ of the inferred machine as D is increased.

Next, let us consider the case where a small amount of white noise is added to the raw data that produced the symbol stream \bar{S} . Table 2 provides the number of states and the entropy rate of

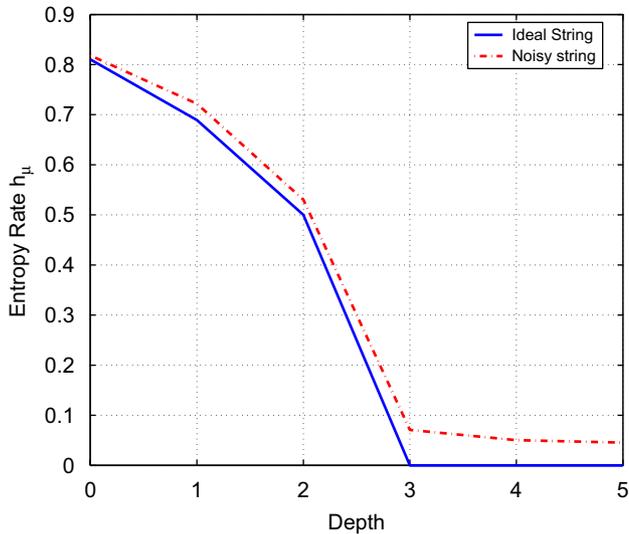


Fig. 4. Entropy rate versus depth.

Table 2
Number of states and entropy rate for noisy string

Depth (D)	No. of states ($\leq \Sigma ^D$)	Entropy rate (h_μ)
0	1	0.818
1	2	0.721
2	4	0.530
3	6	0.070
4	8	0.050
5	12	0.045

the inferred D -Markov machine for various depths. Although the number of inferred states appears to increase with increasing depth, it can be observed that the change in entropy rate h_μ is very small beyond $D = 3$. This means that very little information is gained by increasing the depth and the uncertainty in the system is largely due to the noise. Hence a criterion for the selection of optimal depth of the D -Markov machine can be established in terms of a lower bound on the change in the entropy rate. The curve, shown in Fig. 4 by dashed lines, exhibits the plot of h_μ of the inferred machine as D is increased.

4.2. Measures for quantifying the divergence in patterns

The D -Markov machine, described above, is capable of representing patterns observed in the symbol sequences. In order to quantify the similarity or diversity in the patterns, a measure needs to be defined. This measure is called an anomaly measure \mathbf{M} since it measures the deviations of anomalous patterns from the nominal pattern. The induced norm of the difference between the nominal state transition matrix Π_0 and the state transition matrix for the current pattern Π_k , is a viable candidate for the anomaly measure, i.e., $\mathbf{M}_k = \|\Pi_0 - \Pi_k\|$. Alternatively, measures of anomaly may be derived directly from the state probability vector \mathbf{p} of the D -Markov machine, which is the left eigenvector corresponding to the unique unity

eigenvalue of the (irreducible) Π -matrix. A measure can be defined as $\mathbf{M}_k = \|\mathbf{p}_0 - \mathbf{p}_k\|$ [5,8], where \mathbf{p}_0 and \mathbf{p}_k represent the nominal and the current state probability vectors, respectively. Another candidate for the anomaly measure is the angle between the state probability vectors:

$$\mathbf{M}_k^{ang} = \arccos \left(\frac{\langle \mathbf{p}_0, \mathbf{p}_k \rangle}{\|\mathbf{p}_0\|_2 \|\mathbf{p}_k\|_2} \right), \quad (10)$$

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

The measures, mentioned above, satisfy the requirements for being a metric. But other measures, that do not qualify as a metric, for example, the Kullback–Leibler distance [16] may also be used.

$$\mathbf{M}_k^{kul} = - \sum_{i=1}^{|\Sigma|} \mathbf{p}_k^i \log_2 \frac{\mathbf{p}_k^i}{\mathbf{p}_0^i}. \quad (11)$$

In the experimental analysis described in the next section, both the angle measure and the Kullback measure are utilized.

5. Experimental validation

This section presents the experimental results to validate the concept of *STSA*-based pattern identification on a laboratory apparatus with computer instrumented electronic systems [5,6]. Two nonlinear systems described by Duffing equation and Van der Pol equation [11] are considered. For the Duffing system with exogenous excitation, the motion is chaotic [10] and hence is approximately periodic under quasi-stationary conditions; and the van der Pol system is self-excited and the resulting limit cycle [11] is periodic. Therefore, in both cases, since the trajectories are approximately or exactly periodic, they are confined within a compact region of the phase space.

5.1. Duffing system

The externally excited Duffing equation is a second-order nonautonomous nonlinear differential equation:

$$\frac{d^2 y(t)}{dt^2} + \beta \frac{dy}{dt} + y(t) + y^3(t) = A \cos(\omega t). \quad (12)$$

The amplitude A was equal to 22.0 and $\omega = 5.0$ rad/s. The variables $y^1 = y$ and $y^2 = dy/dt$ constitute the phase space. Behavior of the Duffing system is sensitive to changes in the parameter β . For β in the range of 0.10–0.28, the behavior of the Duffing system in Eq. (12) is largely similar though there are small variations. However, when β increases to ≈ 0.29 , the system undergoes a period doubling bifurcation. The behavior remains essentially unchanged for further increases in β . As evidenced from the phase plots in Fig. 5, there are two distinct patterns. However, there are minor variations in the patterns for $\beta \in [0.1, 0.28]$. As stated earlier, the objective is not merely to classify the patterns correctly, but also to identify the variations in them. Each data set comprised of 2800 data points of the two phase variables y^1 and y^2 sampled at a uniform rate of 100 Hz.

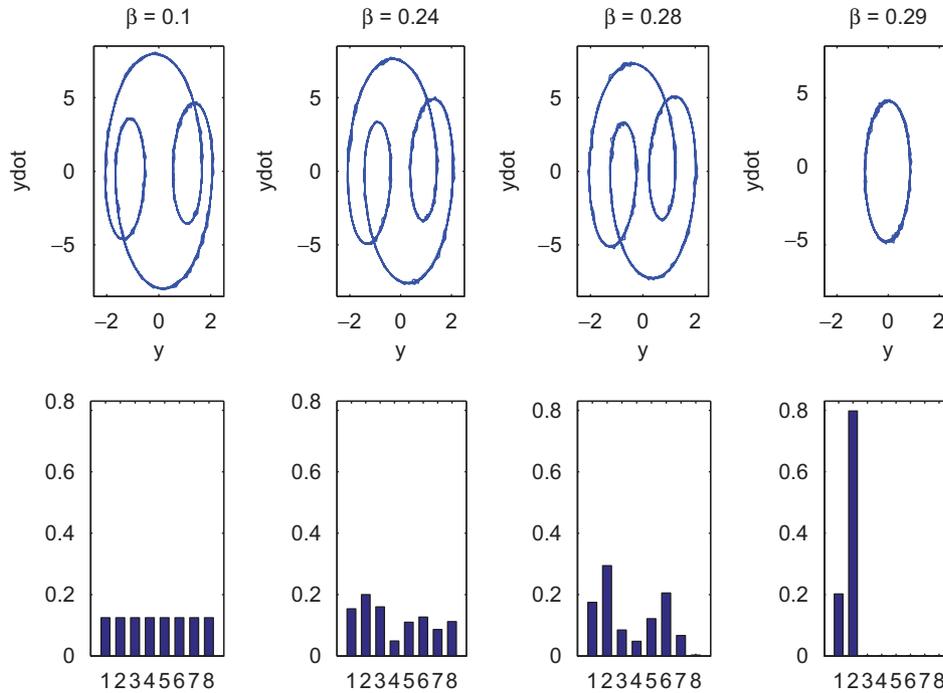


Fig. 5. Duffing system phase plots and state probability histograms.

5.1.1. Generation of partition

The pattern for $\beta=0.10$ is chosen as the nominal since it has the largest span. It is observed in Fig. 5 that the phase trajectory is largely elliptical in nature. Hence, it is appropriate to choose the shape of the partition as an ellipse. The functional for the partitioning boundaries is defined as

$$f(y_i) = \frac{(y_i^1 - \mu_1)^2}{a^2} + \frac{(y_i^2 - \mu_2)^2}{b^2}, \tag{13}$$

where

$$\mu_1 = \frac{1}{N} \sum_{i=1}^N y_i^1 \tag{14}$$

and

$$\mu_2 = \frac{1}{N} \sum_{i=1}^N y_i^2. \tag{15}$$

The semi-minor axis (a) and the semi-major axis (b) are determined from the nominal data as 2.18 and 8.31, respectively. Further, μ_1 is found to be 0.0053 and μ_2 is found to be -0.0597 . The sequence $\{s\}$ is obtained by evaluating the functional for all y_i . The partitioning algorithm in Section 3 yields the alphabet size $|\Sigma| = 8$ with a threshold value $\epsilon_h = 0.05$. The bounds of the partitions are determined from $\{s\}$. The partitions are depicted in Fig. 6. The innermost ellipse and each of the annular regions between successive ellipses represent individual symbols. In general, the partition boundaries are not restricted to be elliptical or ellipsoidal. The partitioning could be a finite

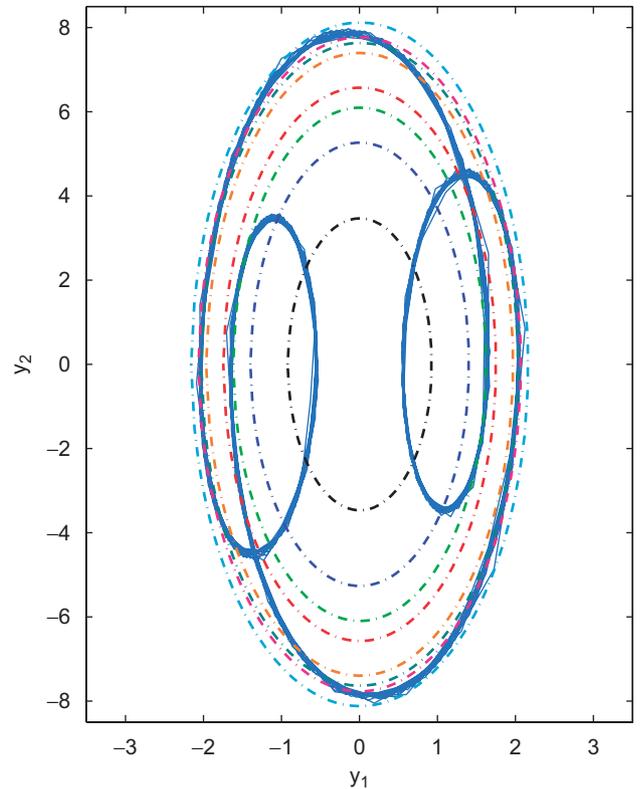


Fig. 6. Partitions for the Duffing system.

set of compact regions in the phase space because the trajectory is confined within a compact region of the phase space as stated in the beginning of this section.

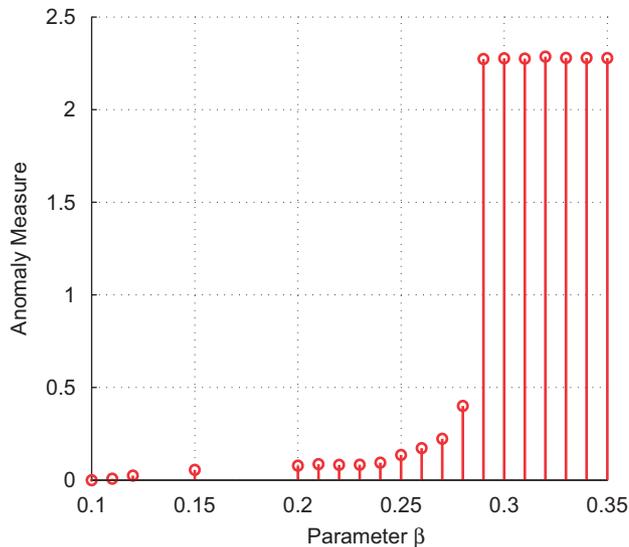


Fig. 7. Kullback measure plot for the Duffing system.

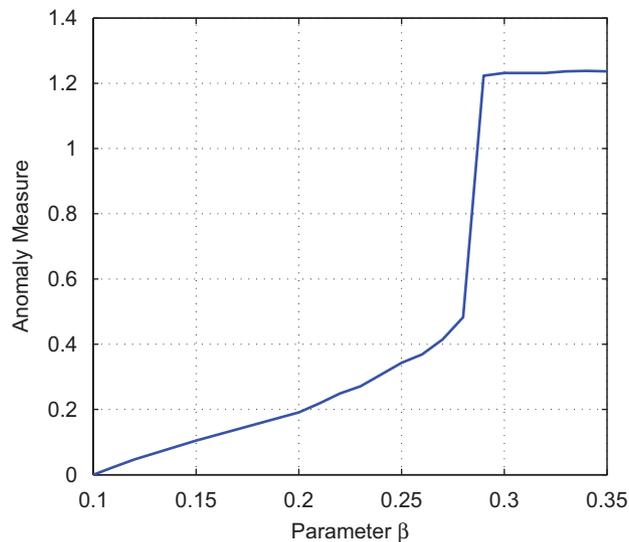


Fig. 8. Angle measure plot for the Duffing system.

5.1.2. Pattern identification

The procedure explained in Section 4 has been utilized to find the optimal depth, which resulted in $D = 1$. Hence, the number of states in the D -Markov machine was eight. Symbol sequences are then generated from all other data sets based on the partition obtained above. As the dynamics of the system change due to variations in β , the statistics of the symbol sequences are also altered and so are the probability distributions that are viewed as patterns. The probability distribution is uniform at the nominal condition of $\beta = 0.10$ because of the ME criterion. As β increases, the distribution deviates from uniform as seen in the histograms of the bottom-row plates of Fig. 5.

The measures defined in Eqs. (10) and (11) have been used for quantifying the changes in the statistics. Fig. 7 depicts the Kullback measure as the parameter β increases from the nominal condition of $\beta = 0.1$. It can be observed from Fig. 7 that the measure for $\beta \in [0.1, 0.28]$ is significantly less ($M < 0.5$) than the measure for $\beta \in [0.29, 0.35]$ ($M \approx 2.3$). Thus, the Kullback measure provides precise classification of the pattern behavior into two distinct categories with no possibility of false classification.

The profile of the angle measure is shown in Fig. 8. This measure also classifies the pattern into two categories even though the rise in measure value is less steeper at $\beta = 0.29$. But more importantly, it identifies the variations in the patterns for β in the range of 0.1–0.28. The angle measure exhibits a gradual monotonic increase for the afore-mentioned values of beta, thereby distinguishing the patterns from one another. It is even able to quantify the variations in patterns for $\beta < 0.15$, which was not possible with the Kullback measure. Thus, by choosing appropriate measures, it is possible to both classify the patterns and identify the variations therein.

The efficacy of $SFNN$ partitioning [15] and ME partitioning, for pattern identification in the Duffing system, are fairly similar. The comparison plot of $SFNN$ partitioning and one-dimensional ME partitioning is provided in Ref. [7]. It can

be observed from Fig. 8 that the results obtained with multi-dimensional ME are comparable with that of $SFNN$ shown in Ref. [7]. However, the computational time required for $SFNN$ partitioning is very high compared to ME partitioning. The time required to generate the partition with $SFNN$ is found to be ≈ 4 h. In the case of one-dimensional ME partitioning, it is found to be ≈ 100 ms and for multi-dimensional ME partitioning it is found to be ≈ 300 ms. Thus, ME partitioning is several orders of magnitude less intensive than $SFNN$ partitioning.

5.2. Van der Pol system

The unforced Van der Pol equation [11] is a second order autonomous nonlinear differential equation

$$\frac{d^2 y(t)}{dt^2} - \mu(1 - y^2(t)) \frac{dy(t)}{dt} + y(t) = 0. \quad (16)$$

Behavior of the Van der Pol system is sensitive to changes in the parameter μ . For small values of μ , the stationary phase trajectory is a smooth orbit, largely similar to a circle of radius 2.0. As μ increases, this shape gets distorted and the distortion is very high around μ equal to 3.0. Though there is no abrupt change in the shape of the phase trajectories, they may be classified into four categories:

1. $\mu \geq 2.0$ —severe distortion,
2. $\mu \in (1.2, 2.0)$ —moderate distortion,
3. $\mu \in (0.5, 1.2)$ —mild distortion,
4. $\mu \leq 0.5$ —little or no distortion.

The phase plots in the top-row plates of Fig. 9 exhibit the phase plots for $\mu = 3.0, 1.4, 0.6$, and 0.2. Each data set comprised of 4500 data points of the variables y^1 and y^2 sampled at 100 Hz.

5.2.1. Obtaining the partition

The data set for $\mu = 3.0$ is chosen as the nominal since it has the largest span. The functional defining the partition is given

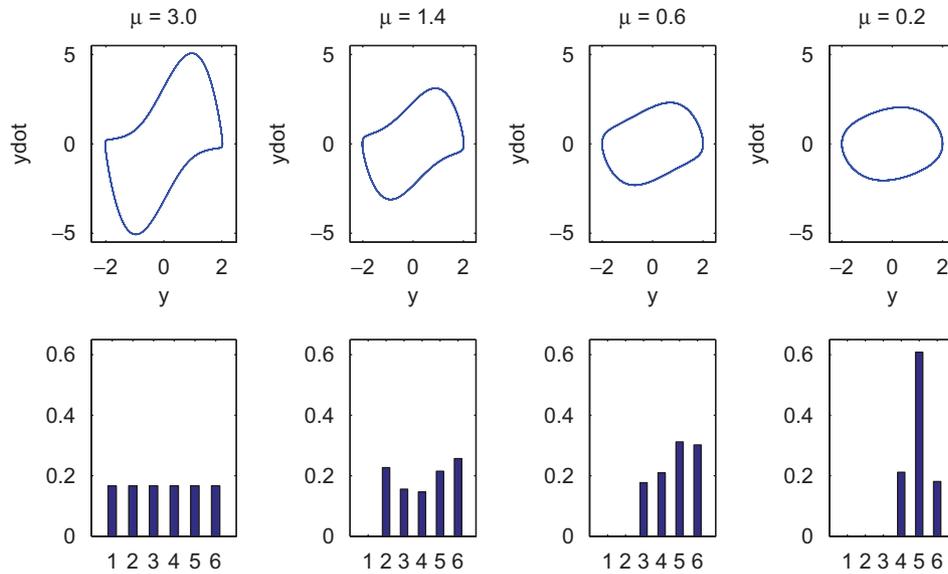


Fig. 9. Van der Pol system phase plots and state probability histograms.

by

$$f(y_i) = \|y_i - c\|_2, \tag{17}$$

where the two-dimensional vector c is defined as

$$c = \frac{1}{N} \sum_{i=1}^N y_i. \tag{18}$$

The value of c in Eq. (18) is found to be [0.0698 0.0548]. The sequence $\{s\}$ is obtained by evaluating the functional for all y_i . The alphabet size $|\Sigma|$ is chosen to be 6 and the depth $D = 1$. The bounds of the partitions are then determined from the sequence $\{s\}$. The partitions are depicted in Fig. 10.

5.2.2. Pattern identification

Symbol sequences are generated from for all data sets, based on the partition defined above. As the dynamics of the system change due to variations in μ , the statistics of the symbol sequences are also altered and so are the state probability distributions that are viewed as patterns. The probability distribution is uniform at $\mu = 3.0$. As μ decreases, the distribution deviates from uniform distribution as seen in the histograms of the bottom-row plates of Fig. 9.

Fig. 11 depicts the Kullback measure as the parameter μ varies from the nominal condition of $\mu = 3.0$. It can be observed from the plot that the measure is approximately zero for category 1 ($\mu > 2.0$) while it varies in the range (0.04–0.2) for category 2 ($\mu \in (1.2, 2.0)$). The measure for category 3 is in the range (0.4–0.6) and for category 4 it is approximately equal to 0.95. Thus, the Kullback measure classifies the patterns into their respective categories accurately.

Fig. 12 provides the angle measure plot. The angle measure shows a gradual increase as μ decreases from its nominal value. This helps in distinguishing patterns even when they

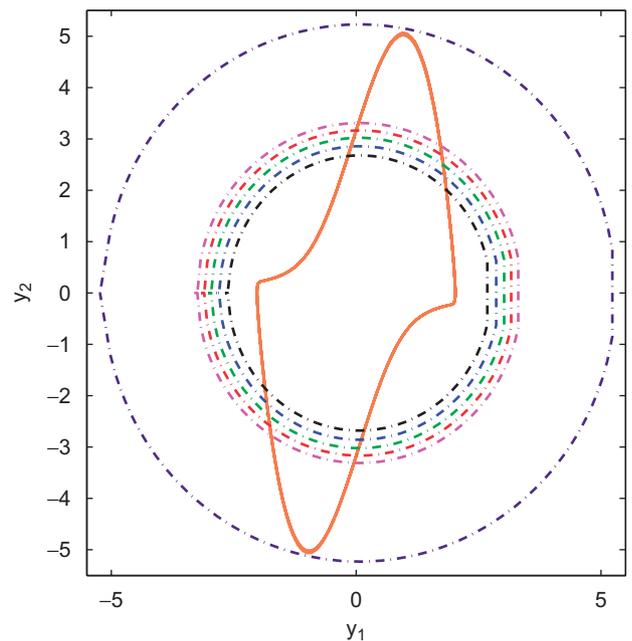


Fig. 10. Partitions for Van der Pol system.

appear apparently the same. For example, the Kullback measure was unable to distinguish between patterns that belong to the severe distortion category. But the angle measure is able to identify their variations as seen by its monotonic increase in that category. Thus, it is possible to both classify the patterns and identify the variations in them by choosing the appropriate measure.

SFNN partitioning was not able to detect patterns in Van der Pol system despite numerous trials. Hence, it is not compared with ME partitioning. The execution time for multi-dimensional ME partitioning was found to be ≈ 200 ms.

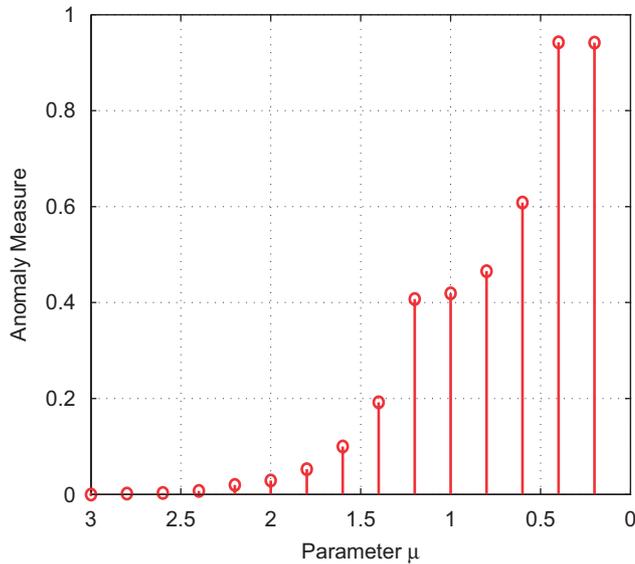


Fig. 11. Kullback distance plot for the Van der Pol system.

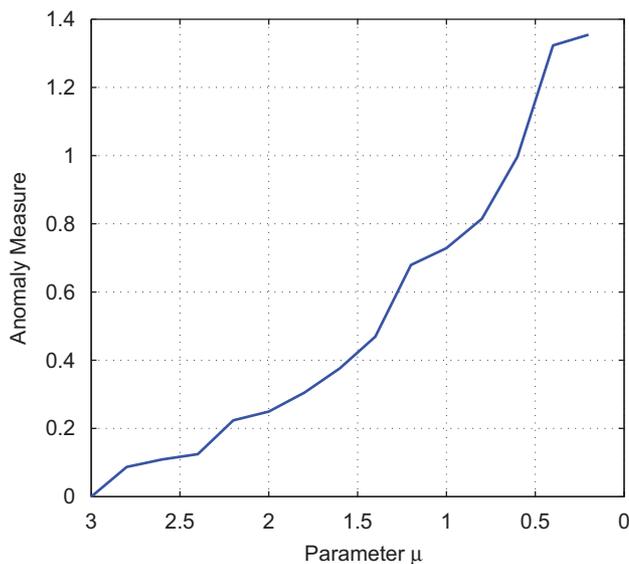


Fig. 12. Angle measure plot for the Van der Pol system.

6. Summary, conclusions and future research

This paper presents a novel method of pattern identification based on symbolic time series analysis (STSA) of multi-dimensional measurement data and its experimental validation on a laboratory apparatus. The time series data are converted into a symbol sequence by partitioning the phase space. The symbol sequences are modeled by finite state automata and patterns are identified from the statistics of the symbol sequences. Specifically, the symbols were obtained with maximum entropy (ME) partitioning. The ME partitioning methodology has been

extended to multiple dimensions in this paper. The symbol sequences are modeled by a special type of *PFSA* with finite memory, known as the *D*-Markov machine. A new procedure for determining the structure of the *D*-Markov, based on entropy rate, is introduced and validated. The efficacy of the proposed technique is demonstrated by validation with experimental data obtained from two nonlinear systems described by the Duffing and the Van der Pol equations.

A major conclusion of this investigation is that pattern identification, based on STSA of multi-dimensional measurement data, can be achieved with both accuracy and computational efficiency. Analytical and experimental research is necessary in the laboratory environment before this tool can be applied to real-life applications. In this context, future research is recommended in the following areas:

- noise reduction in time series for robust pattern identification and
- finding a (possibly) universal design procedure for the partitioning functional.

References

- [1] E. Ott, Chaos in Dynamical Systems, vol. 2/e, Cambridge University Press, UK, 2003 pp. 2–5, 13, 44.
- [2] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. Pattern Anal. Mach. Intell. 22 (1) (2000) 4–37.
- [3] D. Lind, M. Marcus, A Introduction to Symbolic Dynamics and Coding, Cambridge University Press, UK, 1995 p. xi.
- [4] C.S. Daw, C.E.A. Finney, E.R. Tracy, A review of symbolic analysis of experimental data, Rev. Sci. Instrum. 74 (2) (2003) 915–930.
- [5] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, Signal Process. 84 (7) (2004) 1115–1130.
- [6] S. Chin, A. Ray, V. Rajagopalan, Symbolic time series analysis for anomaly detection: a comparative evaluation, Signal Process. 85 (9) (2005) 1859–1868.
- [7] V. Rajagopalan, A. Ray, Symbolic time series analysis via wavelet-based partitioning, Signal Process. 86 (11) (2006) 3309–3320.
- [8] E. Vidal, F. Tollard, C. Higuera, F. Casacuberta, R.C. Carrasco, Probabilistic finite-state machines—part i, IEEE Trans. Anal. Mach. Intell. 27 (7) (2005) 1013–1025.
- [9] R. Duda, P. Hart, D. Stork, Pattern Classification, vol. 2/e, Wiley, New York, 2001 pp. 6–7.
- [10] J.M.T. Thompson, H.B. Stewart, Nonlinear Dynamics and Chaos, Wiley, Chichester, UK, 1986 pp. 2–3.
- [11] M. Vidyasagar, Nonlinear Systems Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1993 pp. 64–65.
- [12] C.J. Veenman, M.J.T. Reinders, E.M. Bolt, E. Baker, A maximum variance cluster algorithm, IEEE Trans. Pattern Anal. Mach. Intell. 24 (9) (2002) 1273–1280.
- [13] T. Chau, A.K.C. Wong, Pattern discovery by residual analysis and recursive partitioning, IEEE Trans. Knowl. Data Eng. 11 (6) (1999) 833–852.
- [14] T.W. Liao, Clustering of time series data—a survey, Pattern Recognition 38 (2005) 1857–1874.
- [15] M.B. Kennel, M. Buhl, Estimating good discrete partitions form observed data: symbolic false nearest neighbors, Phys. Rev. Lett. 91 (2003) 084102.
- [16] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, New York, 1991 pp. 18–19.

About the Author—VENKATESH RAJAGOPALAN earned the Ph.D degree in Electrical Engineering from the Pennsylvania State University in 2007 under the guidance of Prof. Asok Ray. He also has a Master's degree in Electrical Engineering from Penn State. His general research interests include Fault Detection, Pattern Recognition and Signal Processing. Currently, he is working at The MathWorks Inc., Natick, MA.

About the Author—ASOK RAY earned the Ph.D. degree in Mechanical Engineering from Northeastern University, Boston, MA, and also graduate degrees in each discipline of Electrical Engineering, Mathematics, and Computer Science. Dr. Ray joined the Pennsylvania State University in July 1985, and is currently a Distinguished Professor of Mechanical Engineering. Prior to joining Penn State, Dr. Ray held research and academic positions at Massachusetts Institute of Technology and Carnegie-Mellon University as well as research and management positions at GTE Strategic Systems Division, Charles Stark Draper Laboratory, and MITRE Corporation. Dr. Ray has been a Senior Research Fellow at NASA Glenn Research Center under a National Academy of Sciences award. Dr. Ray's research experience and interests include: control and optimization of continuously varying and discrete-event dynamical systems; intelligent instrumentation for real-time distributed systems; and modeling and analysis of complex dynamical systems from thermodynamic perspectives in both deterministic and stochastic settings, as applied to Aeronautics and Astronautics, Undersea Vehicles and Surface Ships, Power and Processing plants, and Robotics. Dr. Ray has authored or co-authored four hundred research publications including about two hundred scholarly articles in refereed journals such as Transactions of ASME, IEEE and AIAA, and research monographs. Dr. Ray is a Fellow of IEEE, a Fellow of ASME, and a Fellow of World Innovative Foundation (WIF).