PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES

Research articles

Spectral invariants of ergodic symbolic systems for pattern recognition and anomaly detection

Najah F. Ghalyan, Chandrachur Bhattacharya, Ibrahim F. Ghalyan and Asok Ray

Published: 20 June 2022 https://doi.org/10.1098/rsta.2021.0196

Abstract

Despite tangible advances in machine learning (ML) over the last few decades, many of the ML techniques still suffer from fundamental issues like overfitting and lack of explainability. These issues mandate requirements for mathematical rigor to ensure robust learning from observed data. In this context, topological invariants in data manifolds provide a rich representation of the underlying dynamical system, which can be utilized for developing a mathematically rigorous ML tool to characterize the dynamical behaviour and operational phases of the underlying process. This paper aims to investigate spectral invariants of symbolic systems for detecting changes in topological characteristics of data manifolds. A novel ML approach is proposed, where commutator norms are used on sequences of endomorphisms to symbolically describe dynamical systems on probability spaces with ergodic measures. The objective here is to detect topological invariants of data manifolds that can be used for signal processing, pattern recognition, and anomaly detection. The proposed ML approach is validated on models of selected chaotic dynamical systems for prompt detection of phase transitions.

This article is part of the theme issue 'Data-driven prediction in dynamical systems'.

1. Introduction

Among the tangible advances in software technologies, graphics processing units (GPUs) have been a great enabler for learning from big data through parallelized computation [1]. This has led to a noticeable success of deep learning to become one of the most important machine learning (ML) methods that have been used in a wide range of disciplines (e.g. engineering [2], medicine [3] and finance [4]). Nevertheless, deep learning suffers from fundamental issues that include overfit [5] and lack of explainability [6], which has led researchers to further investigate alternative ways of *smart learning* from observed data, instead of solely relying on *big data*. In this context, Vapnik & Izmailov [7] have introduced the paradigm of *learning using statistical invariants* (LUSI), which aims at developing a mathematically rigorous approach for regression and pattern recognition by capturing statistical invariants to extract rich information about the underlying stochastic process from observed data. This paradigm relies on weak convergence of risk functionals in the context of reproducing kernel Hilbert spaces [7]; these concepts are intimately related to the topology of data manifolds, generated by the dynamical system of the stochastic process.

Topological data analysis (TDA) is one of the important areas in applied mathematics, which can be gainfully used for development of ML tools to provide an efficient means of information extraction from high-dimensional data in a manner that is insensitive to the selection of a particular metric; it also provides dimensionality reduction and robustness to noise [8]. One of the most important techniques in TDA is the method of persistent homology [9] for clustering and data analysis, which makes use of homology groups of data manifolds to provide information about the underlying dynamical system. For example, in the forced Duffing system [10], a phase transition due to a small change in the dissipation parameter may cause a bifurcation [11], which is associated with the collapse of three homology groups to a single homology group in the data manifold of the phase space, as explained later in the current paper.

The fundamental groups of topological spaces, introduced by Poincaré [12], are the first and simplest homotopy groups [13], and are algebraic invariants that are critically important for characterization and classification of topological spaces [14]. Interestingly, fundamental groups also provide information about the respective covering spaces; in fact, subgroups of the fundamental groups can be used to classify the covering spaces [15]. Furthermore, fundamental groups of a base space can be used to construct covering spaces by using the Lifting theorem [15]. One may think of a base space as the space generated by measurements from m sensors, and a covering space as the space generated by the measurements from these m sensors in combination with additional n sensors. Then, the Lifting theorem can be gainfully used to describe features in the data manifold of the $(m+n)$ sensors by applying the concept of fundamental groups to analyse the data manifold that is generated by the m sensors only. This is important for dimensionality reduction [16], data compression [17] and estimation of dynamical systems with malfunctioning sensor(s) [18].

While TDA techniques, like the persistent homology method, can be efficiently used for characterizing topological invariants by capturing *spatial* patterns in the data manifolds, these topological invariants are often strongly related to the *temporal* and *sequential* behaviour of the underlying dynamical systems which generate these data manifolds. A central theme in the current paper is to develop an ML method, in the setting of symbolic dynamics, that makes use of both spatial and temporal patterns to detect topological invariants of data manifolds, which can be used for downstream analytics (e.g. signal analysis and anomaly detection).

Symbolic dynamics deals with dynamical systems on shift spaces, which consist of semi-infinite and bi-infinite sequences of symbols, defined by a shift-invariant constraint on finite-length data strings [16]. In this context, the concept of symbolic time series analysis (STSA) (e.g. [19,20]) has been used to construct Markov models for anomaly and change detection from observed windows of sensor time-series [11,21]. In the framework of STSA, a (finite-length) time series is partitioned for conversion into a (finite-length) string of symbols from a (finite-cardinality) alphabet $\mathscr{A}$ [22–25]. Subsequently, probabilistic finite state automata (PFSA) are constructed from these symbol strings [21,26,27], in which the probability distribution of the *emitted* symbols depends upon the immediately preceding (at most) D symbols, where D is a positive integer called the Markov depth. Such a PFSA is called a D*-Markov machine*, which has found diverse applications in pattern classification and anomaly detection (e.g. [11,21,28,29]).

In the above setting of a D*-Markov machine*, the selection of the window length, L, of the time series used to construct the PFSA largely depends on the Markov depth D, the alphabet size $|\mathscr{A}|$, and the nature of the particular underlying process that generates the time series [21]. To find a lower bound on the widow-length parameter L, required to estimate the PFSA parameters, one may consider an increasing sequence of L's. Under the assumption of asymptotic statistical stationarity [21], the computed state transition probability matrix converges to a constant matrix; as a consequence, L may become arbitrarily large [11]. Thus, one may choose the smallest L at which the stochastic matrix tends to be approximately time-invariant; the resulting model could be treated as a time-homogeneous Markov chain [30]. However, this scenario would typically require a large value of L, which could be infeasible in many physical applications, where decisions need to be made with low-delay tolerance [31].

The notion of measure-preserving transformation (MPT) has evolved in the discipline of Statistical Mechanics to represent Hamiltonian systems (e.g. Louisville Theorem [32]), where the total energy of the dynamical system is invariant. A key concept in this regard is that, even though a measure-preserving dynamical system is described by a transformation with time-varying eigenvalues, the absolute values of the eigenvectors remain time-invariant under the ergodicity assumption [33]. Based on this rationale, Ghalyan & Ray [34,35] introduced a methodology for constructing PFSA, from short-length windows of time series, which generate non-homogeneous Markov chain models for describing the uncertain dynamics of a physical process. As a result, the time-invariance of the eigenvectors, which reflects measure-invariance and ergodicity of the dynamical system, is used to decide the window length of the time series required to construct the PFSA. Unlike the standard PFSA [11,21], which requires increasing the window length until the resulting PFSA is no longer significantly changing, the framework of MPT-based PFSA [34] can be used to select a minimum window length for which the principal eigenvectors are nearly constant. In this context, a measure of variation in the eigenvectors' absolute values has been used as a metric for an anomaly of the dynamical system [34]. The rationale is that anomalies could often be associated with variations in the system's total energy, which naturally makes the system no longer measure-invariant and thus the eigenvectors are no longer time-invariant. Later, Ghalyan & Ray [35] used the concept of *Cylinders* that represent the topological bases of shift spaces to define a probability measure for the PFSA, produced by symbol strings generated from an ensemble of sensor time-series data. This procedure provides a mathematical foundation for symbol-sequence generation [33] and their relationships with the dynamics of the underlying physical process from a measure-theoretic perspective [35].

The current paper focuses on developing a mathematically rigorous ML method that makes use of topological invariants in data manifolds for pattern recognition and anomaly detection. In the view of this formalism, changes within a given phase are produced by the action of topological (smooth) transformations that preserve topological invariants (e.g. homology groups), while changes between different phases correspond to changes in these topological invariants. Likewise, changes within a cluster or a class are viewed as topological changes that preserve some topological invariants, while changes between different clusters or classes are due to changes in these topological invariants. Relying on this formalism, the paper proposes a symbolic dynamics-based method for detecting topological invariants of data manifolds by investigating spectral invariants of symbolic systems. It is shown that, under the assumptions of ergodicity and measure preservation, commutator norms of a sequence of state transition probability matrices can be efficiently used to predict these topological invariants from the observed time series. This setting is shown to provide a thorough understanding of the application results in dynamic models of three chaotic systems, and lays a concrete mathematical foundation for future work on signal analysis and pattern recognition from observed time series in physical systems such as those reported in [34–36].

Contributions: Major contributions of the paper are summarized below:

(i) *Topological and measure-theoretic analysis of ergodic symbolic systems for ML*: A mathematical framework has been established for ML by relying on the concept of spectral invariants of ergodic symbolic systems.

(ii) *Usage of the commutator norm for ML*: The commutator norm of state transition probability matrices is proposed as a metric for change detection in topological invariants of data manifolds, which can then be used for pattern recognition and anomaly detection.

(iii) *Validation of the proposed methodology on models of chaotic dynamical systems*: The main theme of the reported theoretical innovations is demonstrated on three different standard models of chaotic dynamical systems (e.g. [10,37,38]), which represent a wide range of applications in physics and engineering. In this way, the proposed methodology lays a mathematically rigorous foundation for signal processing, pattern recognition and anomaly detection in uncertain dynamical systems from observed time series.

Organization: The paper is organized in four sections including the present section. Section 2 succinctly presents the mathematical principles of symbolic time series analysis (STSA) and fundamental groups in algebraic topology. This section also develops a measure-theoretic framework for spectral analysis of ergodic symbolic systems, generated from observed time series, which plays a central role in developing the algorithms presented in this paper. Section 3 validates the proposed concept and the underlying algorithms on time-series data, generated from three dynamical models of chaotic systems. Section 4 summarizes and concludes the paper along with recommendations for future research.

2. Introduction to symbolic dynamics and algebraic topology

This section provides the essential mathematical analysis and technical background for modelling dynamical systems in a symbolic setting from an ensemble of observed time series by relying on tools of algebraic topology. The technical approach provides a convenient way to detect changes in topological invariants of data manifolds, generated by dynamical systems. The development of ML methods, which make use of these topological invariants, is important for establishing data-driven models with mathematical rigor so that they are robust to overfit [7]. This is a central theme of the current paper, in which the concept of symbolic dynamics is gainfully used for capturing these topological invariants for signal processing, pattern recognition and anomaly detection in uncertain dynamical systems. While a majority of the details are given in previous publications (e.g. [11,21,35] and references therein), the core concepts are presented below for completeness of the current paper.

(a) Dynamical systems

Let a dynamical system on the probability space $(\Omega, \mathscr{E}, P)$ be described by a quadruple $(\Omega, \mathscr{E}, P, T)$, where the transformation $T : (\Omega, \mathscr{E}, P) \to (\Omega, \mathscr{E}, P)$ is a P-measurable mapping of Ω onto itself.

Definition 2.1.

A measurable set $E \in \mathscr{E}$ is called T-invariant if $P[E \;\Delta\; T^{-1}E] = 0$, which implies that $Tx \in E$ for P-almost all $x \in E$. Furthermore, a function $f : \Omega \to [0, \infty)$ is called T-invariant if $f(Tx) = f(x)$ for P-almost all $x \in \Omega$. A measurable transformation T is called a measure-preserving transformation (MPT) if $P[T^{-1}E] = P[E]\ \forall E \in \mathscr{E}$. (Note: Δ is the symmetric difference on sets such that $A \Delta B \triangleq (A \setminus B) \cup (B \setminus A)$.)

The concept of MPT has been widely used to investigate the asymptotic properties of random sequences in statistical mechanics [33]. For a measure-preserving endomorphism T on a (finite) measure space $(\Omega, \mathscr{E}, P)$, every measurable set E has the recurrence property [32] in the following sense:

Definition 2.2.

A dynamical system $(\Omega, \mathscr{E}, P, T)$ is recurrent if once $E \in \mathscr{E}$ is visited, it would be revisited infinitely many times; that is, if $x \in E$, then there are infinitely many values of n such that $T^{(n)}x \in E$.

Ergodicity is a stronger property than recurrence for dynamical systems, and a formal definition of ergodicity follows.

Definition 2.3 ([33]).

A dynamical system $(\Omega, \mathscr{E}, P, T)$ is said to be ergodic if each $T^{(n)}$-invariant set $E \in \mathscr{E}$ is trivial, i.e. either $P[E] = 0$ or $P[E] = 1 \forall n \in \mathbb{N}$. Equivalently, the measure P is said to be $\{T^{(n)}\}$-ergodic.

Remark 2.4.

The concept of ergodicity has been widely used in Statistical Mechanics and probabilistic modelling of dynamical systems [33]. In an ergodic process, it is sufficient to have a single adequately long realization in order to characterize the statistics of the underlying process. Given a discrete-time realization of an ergodic process as $\{X_n : X_n \in L_1(P)\}$, the time average $\frac{1}{n}\sum_{j=1}^{n} X_j$ converges P-ae and in $L_1(P)$ to the ensemble average $\int_\Omega X_n dP$ as $n \to \infty$ [39].

Definition 2.5 ([40]).

A function $f \in L_1(P)$ is said to be an eigenfunction of a dynamical system $(\Omega, \mathscr{E}, P, T)$ with the eigenvalue $\lambda^{(n)}$ if f is a non-zero function and $f(T^{(n)}) = \lambda^{(n)} f$ P-ae for all $n \in \mathbb{N}$.

Another useful formulation of ergodicity is as follows: Given a probability space $(\Omega, \mathscr{E}, P)$, the sequence of endomorphisms $\{T^{(n)}\}$ is ergodic if and only if every invariant measurable function is a constant P-ae on Ω. Based on this formulation, it can be tested whether a sequence $\{T^{(n)}\}$ is ergodic or not by looking at its eigenfunction f corresponding to the eigenvalue $\lambda^{(n)} = 1$, for which $f(x) = f(T^{(n)}x)$ for P-almost all $x \in \Omega$ and for all $n \in \mathbb{N}$. Hence, f is invariant under $\{T^{(n)}\}$ and therefore is a constant function P-ae if and only if $\{T^{(n)}\}$ is ergodic.

An interesting property stronger than ergodicity is *mixing*, as defined below.

Definition 2.6.

A dynamical system $(\Omega, \mathscr{E}, P, T)$ is said to be mixing if for all sets A and B, the following condition holds:

$$\lim_{n\to\infty} P(T^{-n}A \cap B) = P(A)P(B). \tag{2.1}$$

This means that a moving measurable set A tends to intersect with each fixed measurable set B, and the measure of that part of A which is contained in B is asymptotically proportional to the measure of B. That is, a set A in its motion *mixes* uniformly in the phase space [33], and hence tends to be stochastically independent of each fixed measurable set B as $n \to \infty$ [32].

Remark 2.7.

While mixing is stronger than ergodicity, it only requires asymptotic independence, which is weaker than the iid assumption.

Between ergodicity and mixing, there lies a concept of technical significance, called *weak mixing* [32], as defined below.

Definition 2.8.

A dynamical system $(\Omega, \mathscr{E}, P, T)$ is called weak mixing if, for arbitrary measurable sets, A and B,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} ||P(T^{-k}A \cap B) - P(A)P(B)|| = 0. \tag{2.2}$$

This concept has a strong and surprising influence on the spectral structure of the transformation T, as given by the following theorem.

Theorem 2.9.

A dynamical system$(\Omega, \mathscr{E}, P, T)$*is weak mixing if and only if every eigenfunction*$f(x)$*is equal to a constant for*P*-almost all*$x \in \Omega$.

Proof.

The proof of the theorem is given in [33]. ▪

(b) Measure-invariant symbolic systems

A symbolic representation of the dynamical system $(\Omega, \mathscr{E}, P, T)$ can be generated by partitioning the space $(\Omega, \mathscr{E}, P)$.

Definition 2.10.

A *partition* of the probability space $(\Omega, \mathscr{E}, P)$ is a (non-empty) finite-cardinality family of pairwise disjoint (non-empty) members of $\mathscr{E}$, whose union is Ω.

Given a partition $\alpha = \{A_1, A_2, \ldots, A_{|\alpha|}\}$ of $(\Omega, \mathscr{E}, P)$, an alphabet $\mathscr{A} = \{a_1, a_2, \ldots, a_{|\alpha|}\}$ is constructed, where the symbols a_i's bear a one-to-one correspondence to the members A_i's of the partition α. Let $\mathscr{F}$ be the σ-algebra generated by the alphabet $\mathscr{A}$, and let ν be a probability measure such that $\nu(\{a_i\}) = P(A_i)\ \forall a_i \in \mathscr{A}$, which defines the measure space $(\mathscr{A}, \mathscr{F}, \nu)$.

Let $\mathscr{A}^{\mathbb{N}}$, where $\mathbb{N} \triangleq \{1, 2, 3, \ldots\}$, denote the set of all one-sided semi-infinite symbol sequences, i.e. $\mathscr{A}^{\mathbb{N}} \triangleq \{s_1, s_2, \ldots : s_k \in \mathscr{A} \text{ and } k \in \mathbb{N}\}$. Hence, the transformation $T : (\Omega, \mathscr{E}, P) \to (\Omega, \mathscr{E}, P)$ and a partition α of $(\Omega, \mathscr{E}, P)$ together generate a (left) shift operator $\Sigma : \mathscr{A}^{\mathbb{N}} \to \mathscr{A}^{\mathbb{N}}$ defined as:

$$\Sigma(s_1, s_2, \ldots) = (s_2, s_3, \ldots). \tag{2.3}$$

Definition 2.11.

Let $n, N \in \mathbb{N}$. A cylinder $C^n_{\sigma_1, \ldots, \sigma_N}$, generated by a block of symbols $(\sigma_1, \ldots, \sigma_N)$, where each $\sigma_i \in \mathscr{A}$, is defined to be the collection of all members $\mathbf{S} \in \mathscr{A}^{\mathbb{N}}$ such that the symbol block $(\sigma_1, \ldots, \sigma_N)$ occurs at the location n, i.e.

$$C^n_{\sigma_1, \ldots, \sigma_N} \triangleq \{\mathbf{S} \in \mathscr{A}^{\mathbb{N}} : s_n = \sigma_1, \ldots, s_{n+N-1} = \sigma_N\}, \tag{2.4}$$

and is called a centred cylinder if it has the form $C^1_{\{\sigma_1, \ldots, \sigma_N\}}$.

Remark 2.12.

A cylinder is both an open and a closed subset of $\mathscr{A}^{\mathbb{N}}$ [41]. Moreover, the centred cylinders $C^1_{\{s_1, \ldots, s_N\}}$, $N \in \mathbb{N}$, form a topological basis for $\mathscr{A}^{\mathbb{N}}$ [42].

Let $(\mathscr{A}^{\mathbb{N}}, \mathscr{F}_\Pi)$ denote the Cartesian product of countably infinitely many copies of the measurable space $(\mathscr{A}, \mathscr{F})$, where $\mathscr{F}_\Pi$ is the product σ-algebra generated by the cylinders $C^n_{\sigma_1, \ldots, \sigma_N}$ for all $n, N \in \mathbb{N}$, corresponding to all feasible initial conditions of the dynamical system $(\Omega, \mathscr{E}, P, T)$ [43]. Let m define a probability measure on $\mathscr{F}_\Pi$, given by:

$$m\big(C^n_{\{\sigma_1, \ldots, \sigma_N\}}\big) \triangleq \nu(s_n = \sigma_1, \ldots, s_{n+N-1} = \sigma_N). \tag{2.5}$$

That is, the probability ν of a symbol block is equal to the measure of the cylinder generated by that symbol block. In this way, a probability measure space $(\mathscr{A}^{\mathbb{N}}, \mathscr{F}_{\Pi}, m)$ and a shift system $(\mathscr{A}^{\mathbb{N}}, \mathscr{F}_{\Pi}, m, \Sigma)$ are defined. It follows from equation (2.5) that the symbolic representation of the dynamical system $(\Omega, \mathscr{E}, P, T)$ is stationary if and only if the measure m of the generated cylinders is n-invariant. In this context, a *stationary symbolic system* is the shift system $(\mathscr{A}^{\mathbb{N}}, \mathscr{F}_{\Pi}, m, \Sigma)$ whose cylinders are n-invariant.

A partition α of the probability space of a dynamical system $(\Omega, \mathscr{E}, P, T)$ generates a $(\mathscr{E} - \mathscr{F}_{\Pi})$-measurable coding (also called partitioning [33]) $\Phi^{\alpha} : \Omega \to \mathscr{A}^{\mathbb{N}}$, which is defined as:

$$\left(\Phi^{\alpha}(\omega)\right)_n = a_i \in \mathscr{A} \text{ if and only if } T^{(n)}(\omega) \in A_i \quad \forall \omega \in \Omega \tag{2.6}$$

i.e. the *n*th element of a symbol sequence $\Phi^{\alpha}(\omega)$, where $\omega \in \Omega$, is $a_k \in \mathscr{A}$ if and only if $T^{(n)}(\omega) \in A_k$ and $A_k \in \alpha$.

Let $n \in \mathbb{N}$ and X_n be a stochastic process defined by the dynamical system $(\Omega, \mathscr{E}, P, T)$. Then, $\mathbf{S}_n \triangleq (\Phi^{\alpha}(X))_n$ is a random variable on the probability space $(\mathscr{A}, \mathscr{F}, \nu)$, and $\{\mathbf{S}_n\}$ defines a *symbolic* stochastic process on the shift space $(\mathscr{A}^{\mathbb{N}}, \mathscr{F}_{\Pi}, m)$. Therefore, for every stochastic process $\{X_n\}$, there exists a symbolic representation $\{\mathbf{S}_n\}$ that is identified by finding a partition of the measure space of the original stochastic process.[1] If the transformation T is measure-preserving, the resulting symbolic representation is called a *Measure-Invariant Symbolic System*.

It is noted that the measure-invariance property of a sequence of transformations does not guarantee stationarity of the corresponding symbolic stochastic process. Therefore, the probabilistic finite-state automata (PFSA) (see §2d) generated by partitioning the probability space of the dynamical system $(\Omega, \mathscr{E}, P, T)$ are non-stationary in general. Consequently, the resulting state transition probability matrices of D-Markov machines (see also §2d) are non-homogeneous in general.

(c) Fundamental groups of data manifolds

This subsection provides a brief introduction to fundamental groups of data manifolds, which form the backbone of the statistical pattern classification and anomaly detection methodology, presented in the current paper. While the details are given in standard literature on algebraic topology (e.g. [12,13]), the essential concepts are presented below for completeness of the paper.

Let $f : I \to X$ and $g : I \to X$, where $I = [0, 1]$, be continuous maps represented by two paths in a data manifold X. These two paths are said to be *homotopic*, denoted as $f \simeq g$, if each of these two functions can be continually deformed into the other one. Formally speaking, $f \simeq g$ if there exists a continuous map $H : I \times I \to X$ such that:

$$H(s, 0) = f(s) \quad \text{and} \quad H(s, 1) = g(s) \; \forall s \in I.$$

The map H is called a *homotopy* between f and g. If $f \simeq g$ and both functions have the same initial point and the final point, i.e.

$$H(0,t) = x_0 \quad \text{and} \quad H(1,t) = x_1 \ \forall t \in I,$$

for some $x_0, x_1 \in X$, then f and g are said to be *path homotopic*, which is denoted as $f \simeq_p g$. The concepts of homotopy and path homotopy are illustrated in figure 1*a*,*b*, respectively.

Figure 1. Concepts of (*a*) homotopy, (*b*) path homotopy, (*c*) multiloop classes of homotopy. (Online version in colour.)

Download figure | Open in new tab | Download PowerPoint

If f is a path in X from x_0 to x_1 (i.e. $f(0) = x_0$ and $f(1) = x_1$), and g is a path in X from x_1 to x_2 (i.e. $g(0) = x_1$ and $g(1) = x_2$), then the path product $*$ is a binary operation defined as:

$$f * g(s) = \begin{cases} f(2s), & \text{for } s \in \left[0, \dfrac{1}{2}\right]. \\ g(2s-1), & \text{for } s \in \left[\dfrac{1}{2}, 1\right]. \end{cases} \tag{2.7}$$

A *loop based at* $x_0 \in X$ is a path that begins and ends at the same point x_0. A loop that begins and always stays at x_0 is called a *constant loop*. Figure 1*c* displays the existence of multiple loops in a data manifold X, where f and f' are homotopic but f and g are not. Furthermore, r is homotopic to the constant loop based at x_0. In this case, r is said to be *nulhomotopic*, and r is *contractible to* x_0. Referring to figure 1*c*, although w may not be homotopic to f or g, it can be shown that w is path homotopic to $f * g$ [41]. Further, since $\simeq_p$ is an equivalence relation, it follows that $w \in [f * g]$, where $[f * g]$ is the equivalence class that contains $f * g$. In this context, the notion of fundamental groups is introduced below.

Definition 2.13.

(Fundamental Group) Let X be a topological space. The set of path homotopy classes of loops based at a point $x_0 \in X$, with the operation $*$ of path product, is called the fundamental group of X relative to the base point x_0. It is denoted by $\pi_1(X, x_0)$.

A topological space X is said to be *path connected* if every pair of points of X can be connected by a path in X. For example, the topological spaces in figure 1 are path connected. If X is path connected and if x_0 and x_1 are two points of X, then $\pi_1(X, x_0)$ is isomorphic to $\pi_1(X, x_1)$ [41].

Therefore, it is logical to identify all the fundamental groups $\{\pi_1(X,x):\forall x\in X\}$ and speak simply of the fundamental group of X, written as $\pi_1(X)$.

Remark 2.14.

Fundamental groups are algebraic invariants that remain unchanged under homeomorphisms.[2] They are critically important for understanding geometric properties of data manifolds. In this context, the fundamental group measures 'the number of holes' in a space [13]. For example, the fundamental group of the manifold shown in figure 1*c* is homeomorphic to $\mathbb{Z}^2$, written as $\pi_1(X)\sim\mathbb{Z}^2$. On the other hand, the manifolds in figure 1*a*,*b* have no holes and thus every loop is homotopic to the constant loop. In this case, the manifold is called simply connected, and it has a trivial fundamental group, often indicated as $\pi_1(X)=0$. In low-order dynamical systems, fundamental groups can often be computed by visual inspection through observation of the number of holes in the data manifold.

(d) Probabilistic finite-state automata

As seen in §2b, a partition of the probability space $(\Omega,\mathscr{E},P)$ of a given dynamical system $(\Omega,\mathscr{E},P,T)$ generates a symbol string from a time series, which in turn is used to construct a probabilistic finite state automaton (PFSA). The PFSA model describes the statistics of the underlying stochastic process on the shift space $(\mathscr{A}^{\mathbb{N}},\mathscr{F}_{\Pi},m)$. The probability of the generated PFSA states is given by the measure of the corresponding cylinders in $(\mathscr{A}^{\mathbb{N}},\mathscr{F}_{\Pi},m)$. The following definitions, which are available in standard literature (e.g. [11,21]), are recalled here for completeness of the paper.

Definition 2.15.

A finite-state automaton (FSA) G, having a deterministic algebraic structure, is a triple $(\mathscr{A},\mathscr{Q},\delta)$, where

— $\mathscr{A}$ is a (non-empty) finite alphabet, i.e. $|\mathscr{A}|\in\mathbb{N}$.

— $\mathscr{Q}$ is a (non-empty) finite set of states, i.e. $|\mathscr{Q}|\in\mathbb{N}$.

— $\delta:\mathscr{Q}\times\mathscr{A}\to\mathscr{Q}$ is a state transition map.

Definition 2.16.

A symbol block, also called a word, is a finite-length string of symbols belonging to the alphabet $\mathscr{A}$, where the length of a word $w \triangleq s_1 s_2 \cdots s_\ell$ with each $s_i \in \mathscr{A}$ is $|w| = \ell$, and the length of the empty word ϵ is $|\epsilon| = 0$. The parameters of FSA are extended as:

— The set of all words, constructed from symbols in $\mathscr{A}$ and including the empty word ϵ, is denoted as $\mathscr{A}^\star$.

— The set of all words, whose suffix (respectively, prefix) is the word w, is denoted as $\mathscr{A}^\star w$ (respectively, $w\mathscr{A}^\star$).

— The set of all words of (finite) length ℓ, where $\ell \in \mathbb{N}$, is denoted as $\mathscr{A}^\ell$.

Remark 2.17.

A symbol string (or word) w that occurs at time n generates a cylinder C_w^n in the symbol-sequence space $\mathscr{A}^{\mathbb{N}}$, and the probability of a string w is given by the measure of that cylinder, i.e. $P(w) = m(C_w^n)$ (see §b).

Definition 2.18.

A probabilistic finite state automaton (PFSA), $\mathscr{K}$, is a pair (G, π), where:

— The deterministic FSA, G, is the underlying algebraic structure of the PFSA, $\mathscr{K}$.

— The morph function (also known as the symbol generation probability function) $\pi : \mathscr{Q} \times \mathscr{A} \to [0, 1]$ satisfies the condition: $\sum_{\sigma \in \mathscr{A}} \pi(q, \sigma) = 1$ for all $q \in \mathscr{Q}$.

The state transition probability mass function $\kappa : \mathscr{Q} \times \mathscr{Q} \to [0, 1]$ is constructed by combining δ and π, which can be structured as a $|\mathscr{Q}| \times |\mathscr{Q}|$ matrix $\mathscr{M}$. In that case, the PFSA can be described as the triple $\mathscr{K} = (\mathscr{A}, \mathscr{Q}, \mathscr{M})$.

When constructing a D-Markov machine, the generation of the next symbol is assumed to depend only on a *finite* history of at most D consecutive symbols, i.e. a symbol block not exceeding the specified length D. In this context, a D-Markov machine [11,21] is defined as follows.

Definition 2.19.

A D-Markov machine is a PFSA, in the sense of definition 2.18, which generates symbols that solely depend on the (most recent) history of at most D consecutive symbols, where the positive integer D is called *Markov depth* of the machine. Equivalently, a D-Markov machine is a stochastic process $S = \cdots s_{-1}s_0s_1\cdots$, where the probability of occurrence of a new symbol depends only on the last consecutive (at most) D symbols, i.e.

$$P[s_n \mid \cdots s_{n-D}\cdots s_{n-1}] = P[s_n \mid s_{n-D}\cdots s_{n-1}]. \tag{2.8}$$

Consequently, for $w \in \mathscr{A}^D$ (see definition 2.16), the equivalence class $\mathscr{A}^\star w$ of all (finite-length) words, whose suffix is w, is qualified to be a D-Markov state that is denoted as w.

Remark 2.20.

Let $x_{n+1} = T(x_n)$ be a chaotic map with initial condition $x_0 \in \Omega$, which is randomly selected based on a probability measure P, where $x_0 \sim P$ generates a dynamical system $(\Omega, \mathscr{E}, P, T)$. Then,

(i) The probability of visiting $E \in \mathscr{E}$ at time n is given by $P(T^n x_0 \in E) = P(x_0 \in T^{-n}E)$.

(ii) Partitioning of the state space $(\Omega, \mathscr{E}, P)$ generates a symbolic representation described by a PFSA, $\mathscr{K} \triangleq (\mathscr{A}, \mathscr{Q}, \mathscr{M})$, where $\mathscr{M}$ is the state transition probability matrix which is a symbolic transformation of x_n to x_{n+1} under the action of $T : (\Omega, \mathscr{E}, P) \to (\Omega, \mathscr{E}, P)$ and uncertain $x_0 \in \Omega$ with $x_0 \sim P$. This generates a probabilistic transformation of a state $q_n \in \mathscr{Q}$ into a state $q_{n+1} \in \mathscr{Q}$ through the state transition probability matrix $\mathscr{M}$.

(iii) Let $v^{(n)}$ denote the state probability vector (i.e. the left eigenvector of $\mathscr{M}^{(n)}$ with respect to the (unique) unity eigenvalue) at time n such that $v^{(n+1)} = v^{(n)}\mathscr{M}^{(n)}$. Now, by denoting $\mathscr{P}^{(n)} \triangleq \mathscr{M}^{(0)}\mathscr{M}^{(1)}\cdots\mathscr{M}^{(n)}$, it follows that $v^{(n)} = v^{(0)}\mathscr{P}^{(n)}$.

(iv) Let a sequence of PFSA be defined as $\{\mathscr{K}^{(n)}\} \triangleq \{(\mathscr{A}, \mathscr{Q}, \mathscr{M}^{(n)})\}$. From definition 2.8 and definition 2.18, it follows that $\{\mathscr{K}^{(n)}\}$ is measure-preserving and weak mixing on the probability space $(\mathscr{Q}, \mathscr{E}, P)$ provided that $\mathscr{M}^{(n)}$ is measure-preserving and weak mixing $\forall n \in \mathbb{N}$.

A straightforward result [33], which is central to the current paper, is presented as follows:

Corollary 2.21.

Let$\{\mathscr{K}^{(n)}\}$*be a sequence of PFSA with on a probability space*$(\mathscr{Q}, \mathscr{E}, P)$, *each of which satisfies the measure-preserving and weak-mixing properties. Then, the (sum-normalized) left eigenvector*$v^{(n)}$ *of each*$\mathscr{K}^{(n)}$ *with respect to the (unique) unity eigenvalue is uniformly distributed*, i.e.

$$v^{(n)} = \left[\frac{1}{|\mathscr{Q}|}, \ldots, \frac{1}{|\mathscr{Q}|}\right] \quad \forall n.$$

Now the following claim is made based on the rudimentary principles of linear algebra.

Claim: Let $A, B \in \mathbb{R}^{n \times n}$, where n is a positive integer, serve as linear operators on a dynamical system. Let $\{v^1, \ldots, v^m\}$ be a set of linearly independent $(n \times 1)$ vectors in $\mathbb{R}^n$, where the positive integer $m \leq n$. Let the m-dimensional subspace $V \subseteq \mathbb{R}^n$ be spanned by $\{v^1, \ldots, v^m\}$. If the linearly independent $(n \times 1)$ vectors $\{v^1, \ldots, v^m\}$ serve as common right eigenvectors of the matrix operators A and B, then A and B commute when they are restricted to operate in the subspace V, i.e. $(AB - BA)y = \mathbf{0}\ \forall y \in V$.

Justification of claim: Given that the vectors $\{v^1, \ldots, v^m\}$ serve as common right eigenvectors of the matrix operators A and B, let $\{\mu_1, \ldots, \mu_m\}$ and $\{\nu_1, \ldots, \nu_m\}$ be the respective eigenvalues of A and B, corresponding to the (common) right eigenvectors $\{v^1, \ldots, v^m\}$.

Let $y \in V \setminus \mathbf{0}$ be arbitrary. Since the set $\{v^1, \ldots, v^m\}$ of m linearly independent vectors form a basis of the vector space V, there exist unique scalars $\alpha_1, \ldots, \alpha_m$ such that $y = \sum_{i=1}^m \alpha_i v^i$. Then, it follows that

$$\begin{aligned}(AB - BA)y &= (AB - BA)\sum_{i=1}^m \alpha_i v^i \\ &= \sum_{i=1}^m \alpha_i \left(\mu_i \nu_i - \nu_i \mu_i\right)\ v^i = \mathbf{0}\end{aligned}$$

End of justification of the claim

Corollary 2.22.

(Corollary 1 of the claim) The above claim is also true if the linearly independent $(1 \times n)$ *vectors* $\{(v^1)', \ldots, (v^m)'\}$ *in* $\mathbb{R}^n$ *serve as common left eigenvectors of the matrix operators* A *and* B.

Corollary 2.23.

(Corollary 2 of the claim) If each of the matrices A *and* B *has* n *linearly independent eigenvectors* (i.e. $m = n$) *and if all these* n *eigenvectors are shared by* A *and* B, *then the matrix operators* A *and* B *commute* (*i.e.* $AB = BA$).

Remark 2.24.

The following two observations are made from the above analysis:

(i) From corollaries 2.21 and 2.23, it follows that, for a sequence of measure-preserving and weak-mixing PFSA, the state transition probability matrices tend to commute. Therefore, the norms of commutators of the state transition matrices are expected to be relatively small. In view of the fact that ergodicity is a mild relaxation of weak mixing [32], it is conjectured that the commutator norm of the evolving state transition probability matrices would be relatively small for ergodic and measure-preserving PFSA.

(ii) While the state space of a dynamical system can be partitioned to make the constructed symbolic system be ergodic in the nominal phase, one may encounter cases where the symbolic system is only recurrent (but not necessarily ergodic) in the nominal phase; in those cases, there will be a subset of symbols $\mathscr{B}$, where $\mathscr{B} \subset \mathscr{A}$, that are persistently visited in the nominal phase. However, if the subset $\mathscr{A} \setminus \mathscr{B}$ of symbols is never visited during the nominal phase, the symbolic system restricted to $\mathscr{B}$ could still be ergodic in the nominal phase. In either case, whether the nominal-phase of the symbolic system is ergodic on its own or after restriction, choosing a reference state transition matrix for computing the commutator norm from the nominal phase would make the commutator norm expected to be small in the nominal phase and relatively larger upon occurrence of an anomaly, which may be the consequence of a phase transition.

3. Validation with time series from models of chaotic systems

This section validates the conjecture in part (i) of remark 2.24 on three well-known chaotic dynamical systems, namely, forced Duffing [10], Lorenz attractor [37] and Rössler attractor [38], whose model equations are numerically solved by the fixed-step fourth-order Runge–Kutta integration algorithm. In these chaotic systems, the evidence of bifurcation is observed from the Poincare section [44] in each of the phase plots that are constructed from the respective time series of a state variable x and its time derivative $\mathrm{d}x/\mathrm{d}t$, while a specific parameter of the underlying chaotic system is varied over a given range.

The technical approach for low-latency detection of phase transitions in these chaotic systems relies on using the difference norm ρ_{dif}, commutator norm ρ_{com}, and a moving reference commutator norm $\tilde{\rho}_{\text{com}}$, which are defined on a sequence of evolving state transition probability matrices $\mathscr{M}^{(n)}$ (see part (iii) of remark 2.20) as:

$$\text{difference norm:}\quad \rho_{\text{dif}}^{(n)} \triangleq ||\mathscr{M}^{(n)} - \mathscr{M}^{(0)}|| \tag{3.1}$$

and

$$\text{commutator norm:}\quad \rho_{\text{com}}^{(n)} \triangleq ||\mathscr{M}^{(n)}\mathscr{M}^{(0)} - \mathscr{M}^{(0)}\mathscr{M}^{(n)}||, \tag{3.2}$$

if $\mathscr{M}^{(0)}$ is taken as the reference point when the varying parameter at the instant 0 represents a nominal (or healthy) condition. The intent here is to detect the occurrence of phase transitions. On the other hand, for a prompt change detection, one may continuously replenish the reference point $\mathscr{M}^{(0)}$ by the immediate past value $\mathscr{M}^{(n-1)}$ as:

$$\tilde{\rho}_{\text{com}}^{(n)} \triangleq ||\mathscr{M}^{(n)}\mathscr{M}^{(n-1)} - \mathscr{M}^{(n-1)}\mathscr{M}^{(n)}||. \tag{3.3}$$

Next a comparison of the above three anomaly measures is investigated for detection of phase transitions in the aforementioned three chaotic systems.

(a) Duffing system

The first chaotic system is described by the forced **Duffing equation** [10], representing the dynamics of a nonlinear spring, which is governed by a second-order differential equation as:

$$\ddot{x} + \beta\dot{x} + \alpha_1 x + \alpha_3 x^3 = A\cos \Omega t \quad \text{with initial conditions: } x(0) = 0;\ \dot{x}(0) = 0. \tag{3.4}$$

where the parameters $\alpha_1 = 1.00$, $\alpha_3 = 1.00$, $A = 22.0$ and $\Omega = 5.00$ are held fixed for all simulation runs, and the dissipation parameter β is varied in the range of 0.10–0.50 at an increment of 0.01 for individual simulation runs. It is noted that the forced Duffing equation is a time-varying second-order nonlinear system, in general, which can be treated as a time-invariant third-order nonlinear system if time is considered as the third state [**44**].

The solution to equation (**3.4**) is sensitive to the dissipation parameter β and also to the initial conditions. Figure 2 displays the steady-state solution for increasing values of the dissipation parameter β, where a phase transition of the system occurs due to a small perturbation in β (in the vicinity of $\beta \approx 0.31$), which causes a bifurcation [**11**] as shown in figure 3*a*. Following the concept of fundamental groups in §2c, it is seen in figure 2 that, before occurrence of a phase transition, the Duffing system generates a data manifold with a fundamental group $\pi_1 \sim \mathbb{Z}^3$ for $0.1 \leq \beta \leq 0.31$, which collapses into a fundamental group $\pi_1 \sim Z^1$ for $0.32 \leq \beta \leq 0.50$ after the phase transition. Therefore, the fundamental group of the data manifold generated by the Duffing system represents a topological invariant that can be used for classifying a system's phases as well as for detection of phase transitions, if any.

Figure 2. Recurrence and ergodicity in the Duffing system with uniform partitioning. (*a*) Ergodic symbolic system, (*b*) recurrent symbolic system. (Online version in colour.)

Download figure | **Open in new tab** | **Download PowerPoint**

Figure 3. Performance of different norms for detection of phase transitions in the Duffing system. (*a*) Bifurcation at $\beta \approx 0.31$, (*b*) difference norm, (*c*) commutator norm, (*d*) norm with moving reference. (Online version in colour.)

Download figure | **Open in new tab** | **Download PowerPoint**

While the Duffing system is recurrent for the given initial condition and dissipation parameter β, it may not be ergodic (see definitions 2.2 and 2.3).[3]

However, a simple uniform partition of the phase space, shown in figure 2, generates a symbolic system that is ergodic before the phase transition, where $\pi_1 \sim \mathbb{Z}^3$, and is only recurrent after the phase transition, where $\pi_1 \sim \mathbb{Z}^1$.

Figure 3 shows the bifurcation diagram and presents the performance of three different norms (see equations (**3.1**)–(**3.3**)) for detection of phase transition in the Duffing system in the vicinity of $\beta \approx 0.31$. Figure 3*b*,*c* shows excellent performance of the commutator and difference norms, given by equations (3.1) and (3.2), separating the two phases, where $\pi_1 \sim \mathbb{Z}^3$ in Phase 1 and $\pi_1 \sim \mathbb{Z}^1$ in Phase 2. As a consequence, difference and commutator norms can be gainfully used for detection of a change in the data manifold's fundamental groups due to a phase transition in Duffing system based on a threshold that can be learned from a held-out validation set. Figure 3*d* shows the performance of commutator norm with the reference state transition matrix $\mathscr{M}_0$ being continuously updated with the immediate past value $\mathscr{M}_{n-1}$, which is ideally suited for identifying the instant of change.

The method of persistent homology [13] may also be used for detection of the aforementioned change in the fundamental groups of data manifolds by detecting the presence of *holes* in the manifold, which can be accomplished by examining the spatial patterns in the phase plots of respective dynamical systems. On the other hand, the anomaly metrics given by equations (3.1)–(3.3) would make use of both the spatial and sequential patterns of the time series generated by the dynamical system for detecting changes in the fundamental groups. This information could be important when the time series of sensor data, used for generating the data manifold, are corrupted by noise. In such situations, the holes in the phase plots may become undetectable.

(b) Lorenz attractor

The second chaotic system is **Lorenz attractor**, which is derived by reducing the order of Navier–Stokes equation [37], and is governed by the following three coupled first-order differential equations:

$$\dot{x} = \sigma(y - x); \quad \dot{y} = x(\rho - z) - y; \ \dot{z} = xy - \beta z \text{ with } x(0) = 1; \ y(0) = 1; \ z(0) = 1, \tag{3.5}$$

where the parameters $\sigma = 10.0$ and $\beta = 8/3$ are held fixed for all simulation runs, and the parameter ρ is varied in the range of 154 to 175 at an increment of 0.10 for individual runs.

Figure 4*a*,*b* shows two two-dimensional phase plots, each given by x and $\mathrm{d}x/\mathrm{d}t$ as abscissa and ordinate, respectively. Following the concept of fundamental groups in §2c, the data manifold in figure 4*a* has a fundamental group $\pi_1 \sim \mathbb{Z}^3$, geometrically represented by three holes in the figure. After phase transition, the manifold tends to have a fundamental group $\pi_1 \sim \mathbb{Z}^2$, by filling up the big hole and still retaining the two small ones, as shown in figure 4*b*. A uniform partition, shown in figure 4*a*,*b*, generates a symbolic system that is recurrent for $\pi_1 \sim \mathbb{Z}^3$ and ergodic for $\pi_1 \sim \mathbb{Z}^2$. This shows that, even with a simple partitioning of the phase space, the distinction between the two topological invariants (i.e. $\pi_1 \sim \mathbb{Z}^3$ and $\pi_1 \sim \mathbb{Z}^2$) of the data manifolds becomes quite obvious in the symbolic dynamics setting.

Figure 4. Recurrence and ergodicity in the Lorenz system with uniform partitioning. (*a*) Recurrent symbolic system and (*b*) ergodic symbolic system. (Online version in colour.)

Download figure | Open in new tab | Download PowerPoint

Remark 3.1.

Following figure 4*a*,*b*, the Lorenz attractor may have many separate trajectories that could look like connected regions. The (possibly existing) holes between trajectories, which are very small relative to the big holes, have been ignored in this paper, because their existence is questionable due to the presence of numerical noise. In fact, implementation of persistent homology method is also expected to yield $\pi_1 \sim \mathbb{Z}^3$ for figure 4*a* and $\pi_1 \sim \mathbb{Z}^2$ for figure 4*b*.

Figure 5*a* shows the bifurcation diagram exhibiting the Poincare section in the $(x, \mathrm{d}x/\mathrm{d}t)$ phase space as a function of the parameter ρ of the Lorenz attractor. It is noticed here that a phase transition occurs around $\rho \approx 166$. This change point is clearly detected by using the three norms, given by equations (**3.1**)–(**3.3**), as shown in figure 5*b*–*d*, respectively.

Figure 5. Performance of different norms for detection of phase transitions in the Lorenz system. (*a*) First bifurcation at $\rho \approx 166$, followed by more bifurcations, (*b*) difference norm, (*c*) commutator norm, (*d*) norm with moving reference. (Online version in colour.)

Download figure | Open in new tab | Download PowerPoint

In the nominal phase (i.e. Phase 1), although the symbolic system is recurrent, it is not ergodic; this is so because the only visited set of symbols is a symbolic subset $\mathscr{B}$, i.e. $\mathscr{B} \subsetneq \mathscr{A}$ and the remaining symbols, belonging to $\mathscr{A} \setminus \mathscr{B}$, are not visited in Phase 1. In other words, the symbolic system restricted to $\mathscr{B}$, which is persistently visited, would be ergodic in Phase 1, as pointed out in remark 2.24. Thus, by choosing an appropriate reference state transition matrix for computing the anomaly, it is expected that the metrics in equations (**3.1**)–(**3.3**) would be small in the nominal phase and relatively much larger after phase transition, as demonstrated in figure 5*b*–*d*.

(c) Rössler attractor

The third chaotic system is the **Rössler attractor** [38] that represents chemical reaction kinetics. The system dynamics is governed by the following three coupled first-order differential equations:

$$\dot{x} = -y - z; \quad \dot{y} = x + ay; \;\; \dot{z} = b + z(x - c) \text{ with } x(0) = 1;\; y(0) = 1;\; z(0) = 1, \tag{3.6}$$

where the parameters $b = 2$ and $c = 4$ are held fixed for all simulation runs, and the parameter a is varied in the range of 0.25–0.55 at an increment of 0.005 for individual runs.

Figure 6 shows two phase plots, given by x and $\mathrm{d}x/\mathrm{d}t$ as abscissa and ordinate, respectively. Following the concept of fundamental groups in §2c, the phase plot in figure 6*a* represents the dynamics before a phase transition, where $\pi_1 \sim \mathbb{Z}^1$; and the phase plot after the phase transition is shown in figure 6*b*, where the data manifold tends to be simply connected; in this case, every loop is homotopic to the constant loop, and hence the manifold tends to have a trivial fundamental group [**13**] (i.e. $\pi_1 = 0$) (see remark 2.14). Following remark 3.1 for Lorenz attractor, very small holes between trajectories are ignored for Rössler attractor; this is a reason for postulating that $\pi_1 = 0$ for figure 6*b*.

Figure 6. Recurrence and ergodicity in the Rössler system with uniform partitioning. (*a*) Recurrent symbolic system, (*b*) ergodic symbolic system. (Online version in colour.)

Download figure | Open in new tab | Download PowerPoint

A uniform partition of the phase space makes the symbolic system recurrent before phase transition, and ergodic after phase transition. Figure 7*a* displays the characteristics of bifurcations in the Rössler system, and figure 7*b*–*d* show the performance of different norms given by equations (**3.1**), (**3.2**) and (**3.3**), respectively, for detection of phase transitions, if any. These figures clearly show an excellent performance of the proposed anomaly metrics to detect the aforementioned topological changes. Moreover, the commutator norm shows a better performance for detecting bifurcations within Phase 2, i.e. for $a > 0.31$.

Figure 7. Performance of different norms for detection of phase transitions in the Rössler system. (*a*) First bifurcation at $a \approx 0.31$, followed by more bifurcations, (*b*) difference norm, (*c*) commutator norm, (*d*) norm with moving reference. (Online version in colour.)

Download figure | Open in new tab | Download PowerPoint

4. Summary, conclusion and future work

This paper has proposed a novel ML method in the settings of symbolic dynamics and algebraic topology for robust decision-making from an observed ensemble of time series data. In the proposed approach, pattern recognition and anomaly detection are viewed from topological perspectives, where changes within a phase are described by topological (smooth) transformations that preserve topological invariants (e.g. homology groups), while changes between different phases imply changes in these topological invariants. Likewise, changes within a cluster or class are obtained by the action of topological transformations that preserve some of the topological invariants, while changes in different clusters or classes correspond to changes in these topological invariants. As a result, the tasks of clustering, classification or phase change detection are reduced to the detection of these topological invariants.

If the nominal PFSA is stationary, the state transition probability matrix would be time-invariant, which would make the commutator norm zero. However, this norm is expected to be very small so long as the symbolic system is ergodic even though the generated PFSA may not be stationary. In this case, rather than waiting for too long to generate a time series of sufficient length as required for constructing stationary PFSA that correspond to homogeneous D-Markov machines, one may use short-length time series that can generate non-homogeneous D-Markov machines such that the commutator norm is small under a nominal condition, although the state transition probability matrix is allowed to be time-varying. This property is important for anomaly detection with strict delay tolerance. An example is to avert a system failure, where recovery to the nominal condition is extremely difficult or perhaps practically impossible.

Although a potentially viable alternative approach is the usage of cohomology groups, which may simplify the computation of their dual homology groups [13], the methodology proposed in this paper is also computationally efficient for change detection in these homology groups via usage of D-Markov machines [21] in a symbolic-dynamics setting. This approach lays a foundation, which is both mathematically rigorous and computationally efficient, for signal processing, pattern recognition, and anomaly detection in uncertain dynamical systems from an observed ensemble of time series.

Furthermore, TDA methods, like persistent homology, provide efficient means for detecting topological invariants in data manifolds. However, they may fail to capture temporal and sequential patterns of the underlying dynamical system that generates the data manifold. This issue can potentially make the detection of topological invariants very difficult if the generated data are corrupted by noise, which may distort the observed data manifold. In view of the ML method, proposed in this paper, this issue can be largely mitigated if both the spatial and temporal patterns of the dynamical system are used for detecting these topological invariants, which would make the process of learning from observed data more optimal and robust to measurement noise. A key concept here is that the detection of topological invariants in data manifolds can be largely facilitated by the detection of spectral invariants of an ergodic sequence of endomorphisms that symbolically describe the dynamical system that generates the data manifold. These spectral invariants can be efficiently detected by anomaly metrics, proposed here, given by the commutator norm of the state transition matrices of the generated PFSA. In the nominal phase, this norm tends to be zero or very small, which would significantly increase upon occurrence of an anomaly.

The main theme of the theoretical innovation in this paper is demonstrated for three different types of chaotic systems, which represent a wide range of dynamical systems in physics and engineering applications (e.g. [37]). While there are many areas of theoretical and experimental research to enhance the work reported in this paper, the authors suggest the following topics for future research:

(i) *Extension of the reported analysis to three-dimensional phase spaces from two-dimensional phase spaces*: This paper has analysed two-dimensional phase spaces for illustration of an algebraic topology-based method for ML, although each of the three chaotic dynamical systems is essentially three-dimensional time-invariant. The concept of the first homotopy group (i.e. fundamental group π_1) has been used to detect two-dimensional holes in the respective manifolds. Further research is necessary for investigation of the proposed symbolic dynamics approach for detection of topological invariants in the three-dimensional manifolds of the chaotic maps considered in this paper. In this case, it would be interesting to make a comparison of the proposed approach with the method of persistent homology for detecting changes in π_1 and π_2 for these three-dimensional chaotic systems [13].

(ii) *Investigation of ML problems in high-dimensional chaotic dynamical systems*: Major difficulties are anticipated in the implementation of the proposed algorithms for ML in high-dimensional chaotic dynamical systems. Therefore, identification of potential problems and their solution methods are recommended as topics of future research.

(iii) *Extension to chaotic dynamical systems with different initial conditions*: Response of chaotic systems could be extremely sensitive to change in initial conditions [10,43]. Different initial conditions may result in entirely different types of trajectories and phase plots within the same dynamical system, which can make the analysis challenging. A future work is recommended for extending the methodology established in this paper for handling chaotic dynamical systems with various initial conditions.

(iv) *Comparison of the proposed methodology with standard ML techniques*: A thorough investigation is recommended for comparison of the proposed methodology with standard ML techniques, like various configurations of deep learning [7], standard methods of TDA and hidden Markov models [17,30], for anomaly detection & prediction as well as phase classification from different perspectives (e.g. robustness to over-fitting), based on noisy sensor data in real-life situations.

(v) *Validation of the proposed methodology for anomaly prediction and pattern classification in diverse applications*: The proposed methodology needs to be validated for both detection and prediction of forthcoming anomalies as well as their classification in diverse engineering applications (e.g. fatigue damage detection in polycrystalline alloys [34], and timely detection of thermo-acoustic instabilities in combustion systems [35]).

Data accessibility

This article has no additional data.

Authors' contributions

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration

We declare we have no competing interests.

Funding

The work reported in this paper has been supported in part by the US Air Force Office of Scientific Research (AFOSR) under grant no. FA9550-15-1-0400 in the area of dynamic data-driven application systems (DDDAS), and by the US Army Research Office (grant no. W911NF-20-1-0226; Funder ID: 10.13039/100000183). Any opinions, findings and conclusions in this paper are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

Acknowledgements

The first author would like to thank the Higher Committee for Education Development (HCED) in Iraq for their financial support.

Footnotes

1↵ If a stochastic process is defined over both positive and non-positive times, i.e. $\{X_n : n \in \mathbb{Z}\}$, a coding map $\Phi^\alpha : \Omega \to \mathscr{A}^{\mathbb{Z}}$ can be used to generate a symbolic representation given by the *two-sided shift system*$(\mathscr{A}^{\mathbb{Z}}, \mathscr{F}_\Pi, m, \Sigma)$. In this case, the centred cylinders take the form $C^{-N}_{\{\sigma_{-N},\ldots,\sigma_{-1},\sigma_0,\sigma_1,\ldots,\sigma_N\}}$.

2↵ A homeomorphism is a bijective function f between two topological spaces, where both f and

f^{-1} are continuous in their respective topologies [41].

3↵ The steady-state symbolic Duffing system, for fixed initial condition and some of the system parameters, is recurrent, because once a segment in the phase-space is visited, it will be revisited infinitely many times. However, the system may not visit every segment in the phase-space (as seen in figure 2*b*) and therefore is not ergodic. It is noted that the system in figure 2*a* is ergodic. The symbolic representation could be recurrent or ergodic depending on the way the phase space is partitioned. Figure 2 shows a uniform partition (UP) [22] of the system's phase space, where the generated symbolic system is ergodic for $0.1 \leq \beta \leq 0.31$, and is recurrent (but not ergodic) for $0.32 \leq \beta \leq 0.50$.

Disclaimer

BNY Mellon is the corporate brand of The Bank of New York Mellon Corporation and may be used to reference the corporation as a whole and/or its various subsidiaries generally. This material does not constitute a recommendation by BNY Mellon of any kind. The information herein is not intended to provide tax, legal, investment, accounting, financial or other professional advice on any matter, and should not be used or relied upon as such. The views expressed within this material are those of the contributors and not necessarily those of BNY Mellon. BNY Mellon assumes no direct or consequential liability for any errors in or reliance upon this material.

One contribution of 16 to a theme issue 'Data-driven prediction in dynamical systems'.

Recommend to your library

Contact information

Help

ROYAL SOCIETY PUBLISHING

Our journals

Historical context

Open access

Publishing policies

Permissions

Conferences

Videos

Blog

Manage your account

Terms & conditions

Privacy policy

Cookies

THE ROYAL SOCIETY

About us

Contact us

Fellows

Events

Grants, schemes & awards

Topics & policy

Collections

Venue hire

THE ROYAL SOCIETY

Blog