Signal Processing 149 (2018) 68–81

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Symbolic analysis-based reduced order Markov modeling of time series data☆

Devesh K. Jha[a,b], Nurali Virani[a,c], Jan Reimann[d], Abhishek Srivastav[e], Asok Ray[a,d,*]

[a] Department of Mechanical and Nuclear Engineering, Pennsylvania State University, University Park, PA 16802, USA
[b] Mitsubishi Electric Research Laboratories, Cambridge, MA 02139, USA
[c] Machine Learning Laboratory, GE Global Research Center, Niskayuna, NY, USA
[d] Department of Mathematics, Pennsylvania State University, University Park, PA 16802, USA
[e] Machine Learning Laboratory, GE Global Research Center, San Ramon, CA, USA

ARTICLE INFO

Article history:
Received 24 October 2017
Revised 17 February 2018
Accepted 11 March 2018
Available online 12 March 2018

Keywords:
Symbolic dynamics
Markov modeling
Model order reduction

ABSTRACT

This paper presents reduced-order modeling of time-series data for a special class of Markov models using symbolic dynamics. These models are constructed from the time-series signal by partitioning the data and then inferring a probabilistic finite state automaton (PFSA) from the resulting symbol sequence, capturing a finite history (or memory) of symbol strings. In the proposed approach, the size of the temporal memory of a symbol sequence is estimated from spectral properties of the resulting stochastic matrix corresponding to a first-order Markov model of the symbol sequence. Then, agglomerative hierarchical clustering is used to cluster states of the corresponding full-order Markov model to construct a reduced-order Markov model based on information-theoretic criteria with a non-deterministic algebraic structure; the parameters of the reduced-order model are identified from the original model by making use of a Bayesian inference rule. The model size is inferred using an information-theoretic inspired criteria; the Markov parameters of the reduced-order model are identified from the original model by making use of a Bayesian inference rule. The paper also identifies theoretical bounds on the error induced in the reduced-size model in terms of expected Hamming distance between the sequences generated by the original and final reduced-size models. The proposed concept is elucidated and validated by two examples on different data sets. The first example analyzes a set of time series of pressure oscillations in a swirl-stabilized combustor, where controlled protocols are used to induce flame instabilities. Variations in the complexity of the derived Markov model represent how the system operating condition changes from stable to an unstable combustion regime. The second example is built upon a public data set of NASA's repository for prognosis of rolling-element bearings. It is shown that: (i) even with a small state-space, the reduced-order models are able to achieve comparable performance, and (ii) the proposed approach provides flexibility in the selection of a reduced-order model for data representation and learning.

1. Motivation and introduction

Markov models are widely used as a statistical learning tool for uncertain dynamical systems [1], where, in general, a Markov chain with unobserved states is constructed from the associated temporal data; in this setting, the learning task is to infer the states and the corresponding parameters of the Markov chain. In addition to hidden Markov modeling (HMM), several other techniques have been proposed for Markov modeling of time-series data. For example, in symbolic time series analysis (STSA)-based Markov modeling [2,3], the states of a Markov chain are represented as a collection of words (i.e., symbol blocks, also referred to as memory words) of different lengths, which can be identified from the time-series data on a discrete space with finite cardinality [2–5]. The symbols are created from the continuously varying time-series by projection onto a set with finite cardinality. The learning involved for these models is inference of the hyperparameters of discretization and memory. A common ground among all these modeling

☆ This work is partially supported by the U.S. Air Force Office of Scientific Research under Grant No. FA9550-15-1-0400 dealing with dynamic data-driven application systems (DDDAS). Jan Reimann was partially supported by NSF Grant DMS-1201263. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

* Corresponding author.
E-mail addresses: devesh.dkj@gmail.com (D.K. Jha), nurali.virani88@gmail.com (N. Virani), jan.reimann@psu.edu (J. Reimann), srivastav@ge.com (A. Srivastav), axr2@psu.edu (A. Ray).

https://doi.org/10.1016/j.sigpro.2018.03.004
0165-1684/

tools is that a Markov chain is induced by probabilistic representation of a deterministic finite state automaton (DFSA), called probabilistic finite state automaton (PFSA) [6]. While the PFSA-based inference provides a consistent and deterministic graph for learning the algebraic structure of the model, it is generally not a very compact representation and may lead to a large number of states in the resulting Markov model. To circumvent this problem, attempts have been made to reduce the state-space by merging statistically similar states of the model [3]. However, the problem may still persist because these models are constructed by partitioning the phase space of dynamical systems and the merging states that are statistically similar may lead to algebraic inconsistency. On the other hand, if the states are merged to preserve the algebraic consistency, it may lead to statistical impurity in the resulting models (i.e., states that have different statistics could be merged together). Other approaches for state aggregation in Markov chains could be found in [7–9]; however, these tools do not consider inference of the Markov model from the data, which may not be suitable for analysis of dynamic data-driven application systems (DDDAS) [10].

The state space for Markov models, created by symbolic analysis, may grow exponentially with increase in memory or order of the symbolic sequence. Estimating the right memory is critical for temporal modeling of patterns observed in the sequential data. However, some of the states may be statistically similar and thus merging them could reduce the size of state-space. Several researchers (e.g., [11,12]) have reported reduced-order Markov modeling of time-series data from temporal patterns, where the size of temporal memory of the symbolic data is estimated from the spectral properties of a PFSA and the constraint of deterministic algebraic structure is imposed by the end objective due to this choice of the data representation model.

The current paper proposes to merge the states by removing the constraint of deterministic algebraic properties associated with PFSA, where the states of the Markov chain are now collection of words from its alphabet of length estimated in the last step; this state aggregation induces a non-determinism in the finite state model. The parameters of the reduced-order Markov model are estimated by a Bayesian inference technique from the parameters associated with the higher-order Markov model. The reduced-order model for data representation is selected by using information-theoretic criteria, where a unique stopping point terminates the state-merging procedure. A bound on the distortion of the predictive capability of the models is identified for order reduction of the state-space. The final product is a generative model for the data; however, some of the predictive capabilities of a DFSA could be lost.

Contributions. Reduced-order Markov modeling of time series, presented in this paper, is constructed in a PFSA setting with a nondeterministic algebraic structure. Nondeterminism is induced by merging states of a PFSA with deterministic algebraic structure inferred from discrete sequential data, which in turn allows a compact representation of temporal data. In contrast to the approach reported by Mukherjee and Ray [3], the proposed method relies more strongly on information-theoretic concepts to arrive at a consistent stopping criterion for model selection. The resulting reduced-order model has fewer parameters to estimate, which leads to faster convergence and decision-making. These reduced-order Markov models can be used with streaming data for sequential hypothesis testing for early fault detection [13,14]. In addition, a bound is quantified on degradation in the model's predictive capability due to state-space reduction, based on the Hamming distance between the sequences generated by the original model and the reduced-order model. The algorithms presented in the paper are elucidated and validated on two different examples: (i) time series of pressure oscillations, collected from a swirl-stabilized combustor apparatus [15] to monitor thermo-acoustic instabilities, and (ii) a public data set for prognosis of rolling-element bearings [16]. In addition to the results on Markov modeling, some of the results on pressure instabilities could be of independent interest in the combustion community for machine learning and active control, as discussed below.

Excellent surveys on the current understanding of the mechanisms for the combustion instability phenomena could be found in [17–21]. Active combustion instability control (ACIC) with fuel modulation has proven to be an effective method for suppression of pressure oscillations in combustors [22,23]. Based on the work available in literature, one may conclude that the performance of ACIC is primarily limited by the large delay in the feedback loop and the limited actuator bandwidth [22,23]. Early detection of combustion instability can potentially alleviate the problems due to time delays in the ACIC feedback loop and thus possibly improve the combustion performance [13,24–27]. While the results in these papers are encouraging, there is no interpretation of the expected variations in the data-driven model that represent changes in the operating regimes of the underlying process. In contrast to the work reported in the existing literature, the current paper demonstrates the changes in the complexity of the underlying time-series model for the pressure fluctuations as the system moves to instability. In summary, this paper has presented a concept of structural changes in the underlying stochastic model and pertinent parameters due to combustion instabilities.

Organization. The paper is organized in six sections including the present section. Section 2 succinctly provides the background and mathematical preliminaries on symbolic dynamics and Markov modeling. Section 3 describes the details of the technical approach for inferring a reduced-order Markov model from time series data. Section 4 presents validation of the underlying theoretical concept on the experimental data of pressure oscillations from a laboratory-scale combustion apparatus [15]. Section 5 presents the results of validation of the underlying theoretical concept on a public data set [16,28]. Finally the paper is summarized and concluded in Section 6 along with recommendations for future research.

2. Background and mathematical preliminaries

Symbolic analysis of time-series data is a relatively recent tool, where continuous sensor data are converted to symbol sequences via partitioning of the continuous domain [2,29]. The stationary dynamics of the symbols sequences are then modeled as a probabilistic finite state automaton (PFSA), which is defined as follows:

Definition 2.1 (PFSA)**.** A probabilistic finite state automaton (PFSA) is a tuple $\mathcal{M} = (\mathcal{Q}, \mathcal{A}, \delta, \boldsymbol{M})$ where

- $\mathcal{Q}$ is a finite set of states of the automaton having cardinality $|\mathcal{Q}|$;
- $\mathcal{A}$ is a finite alphabet set of symbols having cardinality $|\mathcal{A}|$;
- $\delta : \mathcal{Q} \times \mathcal{A} \to \mathcal{Q}$ is the state transition function;
- $\boldsymbol{M} : \mathcal{Q} \times \mathcal{A} \to [0, 1]$ is the $|\mathcal{Q}| \times |\mathcal{A}|$ emission matrix (also known morph matrix). The matrix $\boldsymbol{M} = [m_{ij}]$ is row stochastic such that m_{ij} is the probability of generating symbol a_j from state q_i.

Remark 2.1. An alternative representation of a PFSA is $\mathcal{M} = (\mathcal{Q}, \boldsymbol{\Pi})$ where $\boldsymbol{\Pi} : \mathcal{Q} \times \mathcal{Q} \to [0, 1]$ is called the $|\mathcal{Q}| \times |\mathcal{Q}|$ state-transition probability matrix. The matrix $\boldsymbol{\Pi} = [\pi_{ij}]$ is row stochastic and π_{ij} is the probability $\Pr(q_j|q_i)$ of visiting state q_j from state q_i. The stationary state probability $\boldsymbol{p}$ of an irreducible (also called ergodic) PFSA is the sum-normalized eigenvector of $\boldsymbol{\Pi}$ corresponding to its (unique) unity eigenvalue [30].

Remark 2.2. The PFSA defined above has a deterministic algebraic structure which is governed by the transition function δ; thus