Pattern Recognition 44 (2011) 1343–1356

Contents lists available at ScienceDirect

Pattern Recognition

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

Wavelet-based feature extraction using probabilistic finite state automata for pattern classification ☆

Xin Jin, Shalabh Gupta, Kushal Mukherjee, Asok Ray *

Department of Mechanical Engineering, The Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 13 July 2010
Received in revised form
2 November 2010
Accepted 4 December 2010
Available online 16 December 2010

Keywords:
Time series analysis
Symbolic dynamics
Feature extraction
Pattern classification
Probabilistic finite state automata

ABSTRACT

Real-time data-driven pattern classification requires extraction of relevant features from the observed time series as low-dimensional and yet information-rich representations of the underlying dynamics. These low-dimensional features facilitate *in situ* decision-making in diverse applications, such as computer vision, structural health monitoring, and robotics. Wavelet transforms of time series have been widely used for feature extraction owing to their time–frequency localization properties. In this regard, this paper presents a symbolic dynamics-based method to model surface images, generated by wavelet coefficients in the *scale-shift* space. These symbolic dynamics-based models (e.g., probabilistic finite state automata (PFSA)) capture the relevant information, embedded in the sensor data, from the associated Perron-Frobenius operators (i.e., the state-transition probability matrices). The proposed method of pattern classification has been experimentally validated on laboratory apparatuses for two different applications: (i) early detection of evolving damage in polycrystalline alloy structures, and (ii) classification of mobile robots and their motion profiles.

1. Introduction

Tools for real-time data-driven pattern classification facilitate performance monitoring of complex dynamical systems if the physics-based models are either inadequate or unavailable [1]. In this regard, a critical issue is real-time analysis of time series for information compression into low-dimensional feature vectors that capture the relevant information of the underlying dynamics [2–4]. Time series analysis is a challenging task if the data set is voluminous (e.g., collected at a fast sampling rate), high-dimensional, and noise-contaminated. In general, the success of data-driven pattern classification tools depends on the quality of feature extraction from the observed time series. To this end, several feature extraction tools, such as principal component analysis (PCA) [3], independent component analysis (ICA) [5], kernel PCA [6], dynamic time warping [7], derivative time series segment approximation [8], artificial neural networks (ANN) [9], hidden Markov models (HMM) [10], and wavelet transforms [11–13] have been reported in technical literature. Wavelet packet decomposition (WPD) [12] and fast wavelet transform (FWT) [13] have been used for extracting rich problem-specific information from sensor signals. Feature extraction is followed by pattern classification (e.g., using support vector machines (SVM)) [4,14].

Recently, the concepts of Symbolic Dynamics [15] have been used for information extraction from time series in the form of symbol sequences [1,16]. Keller and Lauffer [17] used tools of symbolic dynamics for analysis of time series data to visualize qualitative changes of electroencephalography (EEG) signals related to epileptic activity. Along this line, a real-time data-driven pattern identification tool, called *Symbolic Dynamic Filtering* (SDF) [18,19], has been built upon the concepts of Symbolic Dynamics, Information Theory, and Statistical Mechanics [1]. In the SDF method, time series data are converted to symbol sequences by appropriate partitioning [19]. Subsequently, a probabilistic finite-state automata (PFSA) [18] are constructed from these symbol sequences that capture the underlying system's behavior by means of information compression into the corresponding state-transition probability matrices. SDF-based pattern identification algorithms have been shown by experimental validation in the laboratory environment to yield superior performance over several existing pattern recognition tools (e.g., PCA, ANN, Bayesian Filtering, Particle Filter, Unscented Kalman Filtering, and Kernel Regression Analysis) in terms of early detection of small changes in the statistical characteristics of the observed time series [20].

Partitioning of time series is a crucial step for symbolic representation of sensor signals. To this end, several partitioning techniques have been reported in literature, such as *symbolic false*

☆ This work has been supported by the U.S. Office of Naval Research under Grant No. N00014-09-1-0688, and the U.S. Army Research Laboratory and the U.S. Army Research Office under Grant No. W911NF-07-1-0376. Any opinions, findings and conclusions expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

* Corresponding author. Tel.: +1 814 865 6377.
E-mail addresses: xuj103@psu.edu (X. Jin), szg107@psu.edu (S. Gupta), kum162@psu.edu (K. Mukherjee), axr2@psu.edu (A. Ray).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.12.003

Nomenclature

α	scale in wavelet transform
Δt	sampling interval
ℓ	length of window $\mathcal{W}$
$\mathbf{p}$	feature vector (state probability vector)
$\mathbf{x}_k$	time series data collected at instant k
$\mathcal{C}$	set of class labels
$\mathcal{H}$	coordinate set denoting the scale-shift data points
$\mathcal{M}$	mapping for conversion of time series to symbol sequence
$\mathcal{O}$	reduction of the state set $\mathcal{Q}$
$\mathcal{Q}$	set of all possible states in a window $\mathcal{W} \subset \mathcal{H}$
$\mathcal{R}$	interval spanning the range of wavelet coefficient amplitudes
$\mathcal{S}$	wavelet surface profile
$\mathcal{S}_\Sigma$	map defining symbolization of a wavelet surface profile
$\mathcal{W}$	two-dimensional window of size $\ell \times \ell$ to convert a symbol block to a state
μ	anomaly measure
Ω	compact region in the phase space of the continuously varying dynamical systems
Φ_i	mutually exclusive and exhaustive cells in Ω
Π	state transition matrix
$\psi_{s,\tau}$	wavelet basis function ψ with scale factor s and time shift τ
Σ	symbol alphabet
σ_k	symbol generated by $\mathcal{M}$ from time series $\mathbf{x}_k$
τ_i	slow-time epoch i
$\tilde{f}$	noise-corrupted version of original signal f
B	signal bandwidth
d	divergence measure
F_c	center frequency that has the maximum modulus in the Fourier transform of the wavelet
f_p	pseudo-frequency for scale generation
f_s	sampling frequency
$F_{s,\tau}$	wavelet transform of function $f(t) \in \mathbb{H}$, where $\mathbb{H}$ is a Hilbert space
k	level of noise contaminating the signal f
m	number of most probable symbols
$O(N)$	time complexity of an algorithm to complete a problem of size N
o_i	element of the reduced set $\mathcal{O}$
P	map defining the partitioning of interval $\mathcal{R}$
q	state of a symbol block formed by the window $\mathcal{W}$
w	additive white gaussian noise with zero mean and unit variance

nearest neighbor partitioning (SFNNP) [21], *wavelet-transformed space partitioning* (WTSP) [22], and *analytic signal space partitioning* (ASSP) [23]. In particular, the wavelet transform-based method is well-suited for time–frequency analysis of non-stationary signals, noise attenuation, and reduction of spurious disturbances from the raw time series data without any significant loss of pertinent information [24]. In essence, WTSP is suitable for analyzing the noisy signals, while SFNNP and ASSP may require additional preprocessing of the time series for denoising. However, the wavelet transform of time series introduces two new domain parameters (i.e., scale and shift), thereby generating an image of wavelet coefficients. Thus, the (one-dimensional) time series data is transformed into a (two-dimensional) image of wavelet coefficients. In this context, for SDF-based analysis of wavelet-transformed data, the prior work [18,22] suggested stacking of the wavelet coefficients from multiple scales to represent the two-dimensional scale-shift wavelet domain by a one-dimensional data sequence. However, this procedure of conversion of the two-dimensional domain to one-dimensional is non-unique and potentially lossy depending on the choice of the stacking procedure.

This paper extends the concept of SDF for feature extraction in the (two-dimensional) scale-shift domain of wavelet transform without any need for non-unique conversion to one-dimensional sequence. In addition, the proposed method is potentially applicable for analysis of regular images for feature extraction and pattern classification. From these perspectives, the major contributions of the paper are as follows:

1. Development of an SDF-based feature extraction method for analysis of two-dimensional data (e.g., regular images and wavelet transform of time series in the scale-shift wavelet domain).
2. Experimental validation (in the laboratory environment) of the feature extraction method for pattern classification in two different applications:
   (i) Early detection of damage in structural materials for timely prediction of forthcoming failures.
   (ii) Behavior recognition in mobile robots by identification of their type and motion profiles.

The paper is organized into seven sections including the present one. Section 2 briefly describes the concepts of symbolic dynamic filtering (SDF) and its application to wavelet-transformed data. Section 3 presents the procedure of feature extraction from the symbolized wavelet image by construction of a probabilistic finite state automaton (PFSA). Section 4 describes the pattern classification algorithms. Sections 5 and 6 present the experimental validations on two applications: (i) early detection of failures in polycrystalline alloys and (ii) classification of mobile robots and their motion profiles, respectively. The paper is concluded in Section 7 along with recommendations for future research.

2. Symbolic dynamics and encoding

This section presents the underlying concepts of symbolic dynamic filtering (SDF) for feature extraction from time series of sensor signals. Details of SDF have been reported in previous publications for analysis of (one-dimensional) time series [18,19]. A Statistical Mechanics concept of time series analysis using symbolic dynamics has been presented in [1]. This section briefly reviews the concepts of SDF for self-sufficiency of the paper and then presents the extension for analysis of (two-dimensional) wavelet images for feature extraction. The major steps of the SDF method for feature extraction are delineated as follows:

1. Encoding (possibly nonlinear) system dynamics from observed sensor data (e.g., time series and images) for generation of symbol sequences.
2. Information compression via construction of probabilistic finite state automata (PFSA) from the symbol sequences to generate feature vectors that are representatives of the underlying dynamical system's behavior.

2.1. Review of symbolic dynamics

In the symbolic dynamics literature [15], it is assumed that the observed sensor time series from a dynamical system are represented as a symbol sequence. Let Ω be a compact (i.e., closed