Signal Processing 116 (2015) 13–28

Contents lists available at ScienceDirect

Signal Processing

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

Multivariate time-series analysis and diffusion maps

Wenzhao Lian [a,*], Ronen Talmon [b], Hitten Zaveri [c], Lawrence Carin [a], Ronald Coifman [d]

[a] Department of Electrical & Computer Engineering, Duke University, United States
[b] Department of Electrical Engineering, Technion - Israel Institute of Technology, Israel
[c] School of Medicine, Yale University, United States
[d] Department of Mathematics, Yale University, United States

ARTICLE INFO

Article history:
Received 12 January 2015
Received in revised form
2 April 2015
Accepted 3 April 2015
Available online 13 April 2015

Keywords:
Diffusion maps
Bayesian generative models
Dimension reduction
Time series
Statistical manifold

ABSTRACT

Dimensionality reduction in multivariate time series analysis has broad applications, ranging from financial data analysis to biomedical research. However, high levels of ambient noise and various interferences result in nonstationary signals, which may lead to inefficient performance of conventional methods. In this paper, we propose a nonlinear dimensionality reduction framework using diffusion maps on a learned statistical manifold, which gives rise to the construction of a low-dimensional representation of the high-dimensional nonstationary time series. We show that diffusion maps, with affinity kernels based on the Kullback–Leibler divergence between the *local statistics* of samples, allow for efficient approximation of pairwise geodesic distances. To construct the statistical manifold, we estimate time-evolving parametric distributions by designing a family of Bayesian generative models. The proposed framework can be applied to problems in which the time-evolving distributions (of temporally localized data), rather than the samples themselves, are driven by a low-dimensional underlying process. We provide efficient parameter estimation and dimensionality reduction methodologies, and apply them to two applications: music analysis and epileptic-seizure prediction.

1. Introduction

In the study of high-dimensional data, it is often of interest to embed the high-dimensional observations in a low-dimensional space, where hidden parameters may be discovered, noise suppressed, and interesting and significant structures revealed. Due to high dimensionality and nonlinearity in many real-world applications, nonlinear dimensionality reduction techniques have become increasingly popular [1–3]. These manifold-learning algorithms build data-driven models, organizing data samples according to local affinities on a low-dimensional manifold. Such methods have broad applications to, for example, analysis of financial data, computer vision, hyperspectral imaging, and biomedical engineering [4–6].

The notion of dimensionality reduction is useful in multivariate time series analysis. In the corresponding low-dimensional space, hidden states may be revealed, change points detected, and temporal trajectories visualized [7–10]. Recently, various nonlinear dimensionality reduction techniques have been extended to time series, including spatio-temporal Isomap [11] and temporal Laplacian eigenmap [12]. In these methods, besides local affinities in the space of the data, available temporal covariate information is incorporated, leading to significant improvements in discovering the latent states of the series.

The basic assumption in dimensionality reduction is that the observed *data samples* do not fill the ambient space uniformly, but rather lie on a low-dimensional manifold.

* Corresponding author.

http://dx.doi.org/10.1016/j.sigpro.2015.04.003
0165-1684/

Such an assumption does not hold for many types of signals, for example, data with high levels of noise [4, 13–15]. In [14,15], the authors consider a different, relaxed dimensionality reduction problem on the domain of the underlying probability distributions. The main idea is that the varying *distributions*, rather than the samples themselves, are driven by few underlying controlling processes, yielding a low-dimensional smooth manifold in the domain of the distribution parameters. An information-geometric dimensionality reduction (IGDR) approach is then applied to obtain an embedding of high-dimensional data using Isomap [1], thereby preserving the geodesic distances on the manifold *of distributions*.

Two practical problems arise in these methods, limiting their applicability to time series analysis. First, in [14,15] multiple datasets were assumed to be available, where the data in each set drawn from the same distributional form, with fixed distribution parameters. Then, the embedding was inferred in the space of the distribution parameters. By taking into account the time dependency in the evolution of the distribution parameters from a single time series, we may substantially reduce the number of required datasets. A second limitation of previous work concerns how geodesic distances were computed. In [14,15] the approximation of the geodesic distance between all pairs of samples was computed using a step-by-step walk on the manifold, requiring $\mathcal{O}(N^3)$ operations, which may be intractable for large N.

In this paper, we present a dimensionality-reduction approach using diffusion maps for nonstationary high-dimensional time series, which addresses the above shortcomings. Diffusion maps constitute an effective data-driven method to uncover the low-dimensional manifold, and provide a parametrization of the underlying process [16]. The main idea in diffusion maps resides in aggregating local connections between samples into a global parameterization via a kernel. Many kernels implicitly induce a mixture of local statistical models in the domain of the measurements. In particular, it is shown that using distributional information outperforms using sample information when the distributions are available [14]. We exploit this assumption and articulate that the observed multivariate time series $\boldsymbol{X}_t \in \mathbb{R}^N, t=1,\ldots,T$, is generated from a smoothly varying parametric distribution $p(\boldsymbol{X}_t|\boldsymbol{\beta}_t)$, where $\boldsymbol{\beta}_t$ is a local parameterization of the time evolving distribution. We propose to construct a Bayesian generative model with constraints on $\boldsymbol{\beta}_t$, and use Markov Chain Monte Carlo (MCMC) to estimate $\boldsymbol{\beta}_t$. Diffusion maps are then applied to reveal the statistical manifold (of the estimated distributions) using a kernel with the Kullback–Leibler (KL) divergence as the distance measure. Noting that the parametric form of distributions significantly affects the structure of the mapped data, the Bayesian generative model should avoid using a strong informative prior without substantial evidence.

Diffusion maps rely on the construction of a Laplace operator, whose eigenvectors approximate the eigenfunctions of the backward Fokker–Planck operator. These eigenfunctions describe the dynamics of the system [17]. Hence, the trajectories embedded in the coordinate system formulated by the principal eigenvectors of the Laplace operator can be regarded as a representation of the underlying controlling process $\boldsymbol{\theta}_t$ of the time series $\boldsymbol{X}_t$.

One of the main benefits of embedding the time series samples into a low-dimensional domain is the ability to define meaningful distances. In particular, diffusion-maps embody the property that the Euclidean distance between the samples in the embedding domain corresponds to a diffusion distance in the distribution domain. Diffusion distance measures the similarity between two samples according to their connectivity on the low-dimensional manifold [3] and has a close connection to the geodesic distance. Thus, diffusion maps circumvent the step-by-step walk on the manifold [14], computing an approximation to the geodesic distance in a single low-cost operation. Another practical advantage of the proposed method is that we may first reveal the low-dimensional coordinate system based on reference data, and then in an online manner extend the model to newly acquired data with a low computational cost. This is demonstrated further when considering applications in Section 4.

The proposed framework is applied to two applications in which the data are best characterized by temporally evolving local statistics, rather than based on measures directly applied to the data itself: music analysis and epileptic seizure prediction based on intracranial electroencephalography (icEEG) recordings. In the first application, we show that using the proposed approach, we can uncover the key underlying processes: human voice and instrumental sounds. In particular, we exploit the efficient computation of diffusion distances to obtain intra-piece similarity measures on well-known music, which are compared with the state-of-the-art techniques.

In the second application, one goal is to map the recordings to the unknown underlying "brain activity states". This is especially crucial in epileptic seizure prediction, where preseizure (dangerous) states can be distinguished from interictal (safe) states, so that patients can be warned prior to seizures [18]. In this application, the observed time series is the icEEG recordings and the underlying process is the brain state, *e.g.*, preseizure or interictal. IcEEG recordings tend to be noisy, and hence, the mapping between the state of the patient's brain and the available measurements is not deterministic, and the measurements do not lie on a smooth manifold. Thus, the intermediate step of mapping the observations to a time-evolving parametric family of distributions is essential to overcome this challenge. We use the proposed approach to infer a parameterization of the signal, viewed as a model summarizing the signal's distributional information. Based on the inferred parameterization, we show that preseizure state intervals can be distinguished from interictal state intervals. In particular, we show the possibility of predicting seizures by visualization and simple detection algorithms, tested on an anonymous patient.

This paper makes three principal contributions. First, we present a data-driven method to fit flexible statistical models adapted to time series. In particular, we propose a class of Bayesian models with various prior specifications to learn the time-evolving statistics from a *single* trajectory realization that accurately models *local* distributional parameters $\boldsymbol{\beta}$. Second, we complement the analysis with diffusion maps based on the distributional information embodied in time-series dynamics. By relying on a kernel,