Expert Systems With Applications 99 (2018) 25–31

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Multi-scale Gaussian process experts for dynamic evolution prediction of complex systems

Changqing Cheng

Department of Systems Science and Industrial Engineering, State University of New York at Binghamton, Binghamton, NY 13902, United States

ARTICLE INFO

Article history:
Received 20 July 2017
Revised 18 December 2017
Accepted 13 January 2018

Keywords:
Multi-scale Gaussian process
Intrinsic time-scale decomposition
Nonlinear
Nonstationary
Multi-step forecasting

ABSTRACT

Predictive analytics has become an important topic in expert and intelligent systems, with broad applications across various engineering and business domains, such as the prediction of exchange rate in finance, weather and demand for energy using mixture of experts. However, selection of the number of experts and assignment of the input to individual experts remain elusive, especially for highly nonlinear and nonstationary systems. This paper presents a novel mixture of experts, namely, nonparametric multi-scale Gaussian process (MGP) experts to predict the dynamic evolution of such complex systems. Concretely, intrinsic time-scale decomposition is first used to iteratively decompose the time series generated from such complex systems into a series of proper rotation components and a baseline trend component. Those components delineate the true time-frequency-energy patterns of the complex systems at different granularity. A Gaussian process (GP) expert is then applied on each component to predict the system evolution at each scale. MGP circumvent the tedious selection and assignment problems via the nonparametric ITD. Summation of those individual forecasts represents the overall evolution of the original time series. Case studies using synthetic and real-world data elucidated that the proposed MGP model significantly outperforms conventional autoregressive models, composite GP model, and support vector regression in terms of prediction accuracy, and it is particularly effective for multi-step forecasting.

Published by Elsevier Ltd.

1. Introduction

Predictive analytics is increasingly becoming an integral component of the fully functioning expert and intelligent systems with applications across various engineering and business domains, such as prediction for exchange rate and climate using expert systems (Yuksel, Wilson, & Gader, 2012). More specifically, as the high cost associated with additive manufacturing has suppressed the wide applications of the 3D printed products, the AM industries are endeavoring to adopt predictive analytics to prevent as opposed to mitigate quality loss (Cheng, Bukkapatnam, Raff, & Komanduri, 2012). The P4 medicine scheme (Predictive, Preventive, Personalized, and Participatory) has emphasized preventive healthcare instead of reactive disease control for optimum decision making. In addition, predictive analytics has been employed to predict the throughput rate of an automotive assembly line (Bukkapatnam & Cheng, 2010) for resource allocation.

Remarkably, during the past decades, the growth of cloud computing and spiral development of Internet of Things and inexpensive sensors, as well as the widespread use of smart devices have enabled the acquisition of a vast array of data ("big data") and dramatically lowered the costs of information storage and retrieval. This has boosted the application of analytics, and potentially transformed expert systems from knowledge-based to data-driven. Particularly, the temporally varying data, or time series, have garnered enormous attention in the machine learning as well as expert systems communities (Cheng et al., 2015; Lin, Wang, Xie, & Zhong, 2017). Those time series data contain consequential causal and dynamic information about the underlying complex systems or processes, and they enable the expert system to harness fundamental patterns for monitoring, prognosis and decision making purposes.

However, temporally evolving data have posed immerse challenges to the pattern recognition and, consequently, the forecasting and control of complex systems, as those data typically exhibit combined nonlinear and nonstationary characteristics. That said, patterns or relationship and trend in the historical observations are ambiguous or difficult to quantify. Although mixture of experts has been studied to handle those complex systems (Yuksel et al., 2012), it still remains an elusive task to select the number of experts and assign inputs to the individual experts.

In this paper, we present a novel generative mixture of experts model by integrating two nonparametric models, Gaussian process (GP) and intrinsic time-scale decomposition (ITD). Although, as a nonparametric approach, GP largely reduces modeling efforts, the squared exponential covariance function form has confined GP only

E-mail address: ccheng@binghamton.edu

https://doi.org/10.1016/j.eswa.2018.01.021
0957-4174/Published by Elsevier Ltd.

to stationary cases. ITD (Frei & Osorio, 2007) handles nonstationarity of the original time series by decomposing it into a number of proper rotations (see Section 3 for details) at progressively decreasing granularity and a trend baseline. It is nonparametric in that no basis functions are required for the decomposition *a priori*. Further, the number of ITD components fully determines the number of experts. The GP model (individual expert) built upon each component (hence, the name multi-scale GP) captures the covariance structures across different scales. The focus of this paper is on the characterization and forecasting of univariate time series.

The rest of the paper is organized as follows: Section 2 provides a brief review for the classical forecasting models of univariate time series. Section 3 presents the theoretical foundation of multi-scale Gaussian process (MGP). In Section 4, both synthetic and real-world case studies are used to illustrate the effectiveness of the proposed MGP model. Final conclusion remarks are given in Section 5.

2. Research background

In this section, a brief review of the forecasting models for nonlinear and nonstationary time series is provided. For detailed description and comparison of different types of models, the readers may refer to our recent work (Cheng et al., 2015) and also (Lin et al., 2017).

A myriad of forecasting approaches has been investigated in the univariate time series forecasting, including auto-regressive models, neural networks, kernel-based methods, etc. In particular, autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) are the most frequently used classical time series models. ARMA has a simple structure based on the dependence of the variables on the lagged history. It specifies the current observation as a weighted sum of past realizations and error terms. Nonetheless, it only fits stationary or weakly stationary processes, and the weight estimate does not converge in the case of high nonstationarity. ARIMA is an extension to ARMA, and addresses the nonstationary issues by differencing the original time series until it is stationary, yet only limited to simple form of nonstationarity (e.g., trend). Oftentimes, real-world time series manifest multiple nonstationary patterns, including seasonality and random spike, such as the electricity load time series shown in Fig. 7. Here, the electricity load time series profile depicts seasonality at different scales, in terms of demand similarity between day-to-day, week-to-week and year-to-year variations. Multiplicative seasonality ARIMA (Taylor, de Menezes, & McSharry, 2006) and double seasonal Holt–Winters exponential smoothing (Taylor, 2003) have been investigated for such load data. But they failed to represent the nonlinear and nonstationary patterns of the load profile beyond seasonality (Cheng et al., 2015). Therefore, the forecasting accuracy is severely depressed with the above-mentioned models.

In expert and intelligent system, grey system models were studied extensively for time series prediction (Kayacan, Ulutas, & Kaynak, 2010). Those models are based only on a set of the most recent data depending on the size of a sliding window of the predictor. That said, the recent observations are assigned a heavier weight to predict the future value. However, the selection of window size and weight remains a challenging task. Mixture of experts models are also popular for nonstationary time series predictions (Yuksel et al., 2012). However, the selection of expert numbers and assignment of inputs to individual experts oftentimes incur tremendous efforts.

Notably, decomposition-based forecasting models have garnered enormous attention of late. Those models decompose the time series into a number of components at different frequency scales, each of which has significantly reduced complexity compared to the original time series. The widely known Fourier

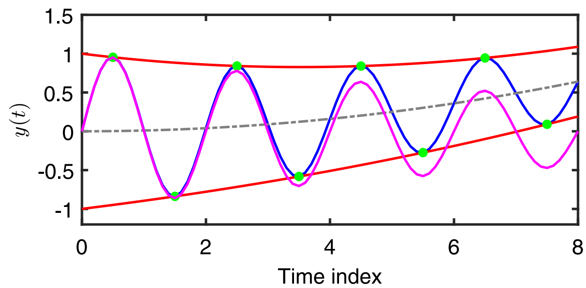

Fig. 1. Schematic diagram of EMD algorithm.

analysis, wavelet analysis, empirical mode decomposition (EMD) (Huang et al., 1998), and the recently developed intrinsic timescale decomposition (ITD) (Frei & Osorio, 2007) are of this category. Fourier analysis describes time series as a sum of sine/cosine basis waves. It is a "global" method and does not capture frequency variation at different temporal locations. Attempt to address this issue typically involves the Windowed Fourier analysis, such as short-time Fourier transform. That is, the long time series is divided into short segments of equal length, and Fourier analysis is then conducted on each segment to represent the local frequency. However, fixed window size is often used; thus high time and frequency resolution cannot be obtained simultaneously. Alternatively, wavelet analysis offers critical time-frequency information across multi-scales and reveals "local" patterns via translation and dilation of basis functions with compact support. Note that both Fourier and wavelet methods require basis functions to be defined *a priori*, which largely determines the effectiveness of these models.

In contrast, EMD (Huang et al., 2003) is nonparametric and does not require any basis function, suitable for any kind of nonlinear and nonstationary patterns. EMD iteratively decomposes the original time series into intrinsic mode functions (IMFs) of progressively lower frequency and a baseline trend. Here, the baseline is defined as the mean of the upper and lower envelopes of the wave in a fixed time window, as indicated in Fig. 1. The upper and lower envelopes are the cubic spline fitting of the local maxima and minima (green dots), respectively. It is evident that the baseline demonstrated the global trend of the time series. When de-trended (i.e., baseline trend component removed), the residual time series, also called IMF (the magenta line), contain the highest frequency presented in the wave at the current level. The low-frequency baseline is further decomposed into another baseline trend and IMF. This process iterates until a baseline trend component without any riding wave is left. Therefrom, the extracted trend baseline represents the general behavior of the time series, and each IMF is modulated in both amplitude and frequency to capture local characteristics.

For example, a forecasting model combining EMD and neural network (NN) has been adopted to predict the exchange rate (Das, Bisoi, & Dash, 2017) with pronounced accuracy compared to NN models. The power of EMD resides in the IMFs with the well-defined instantaneous phase, amplitude and frequency, or time-frequency-energy patterns according to the Hilbert-Huang transform (Frei & Osorio, 2007). In other words, the well-behaved IMFs are proper rotations that are strictly positive at local maxima and strictly negative at local minima. However, IMFs generated from EMD often do not have the integral properties of proper rotations. Later, a sifting process (Huang et al., 1998) was developed to select the best baseline among a pool of candidates and generate the true proper rotation at each decomposition level. Nonetheless, by smoothing out the uneven waves, this procedure often blurs the paramount transient and nonstationary features. To this end,

Download English Version:

https://daneshyari.com/en/article/6855117

Download Persian Version:

https://daneshyari.com/article/6855117

Daneshyari.com