Chemical Engineering Science 123 (2015) 125–136

Contents lists available at ScienceDirect

Chemical Engineering Science

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

Self-active and recursively selective Gaussian process models for nonlinear distributed parameter systems

Pei Sun [a], Junghui Chen [b,*], Lei Xie [a,*]

[a] Institute of Cyber-Systems and Control, National Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, Zhejiang, China
[b] Department of Chemical Engineering, Chung-Yuan Christian University, Chung-Li 320, Taiwan, ROC

HIGHLIGHTS

- GP model is used to model the spatiotemporal process with the assistance of KL decomposition.
- The process outputs are predicted along with the uncertainty of the spatiotemporal process.
- Active data are selected automatically based on the variance to enhance the model.
- The KL–GP model is updated recursively when new information is added.

ARTICLE INFO

Article history:
Received 4 May 2014
Received in revised form
2 September 2014
Accepted 15 October 2014
Available online 22 October 2014

Keywords:
Chemical processes
Gaussian process model
Karhunen–Loève decomposition
Nonlinear dynamics
Spatial-temporal process
Systems engineering

ABSTRACT

Modeling a nonlinear distributed parameter system (DPS) is difficult because it is usually hard to obtain the first-principle models in DPS with strong spatiotemporal characteristics. In this paper, a novel data-driven model, called KL–GP, is proposed based on Karhunen–Loève (KL) decomposition and Gaussian process (GP) models. First, KL decomposition is employed for the time/space separation and dimension reduction. The spatiotemporal output is projected onto a low-dimensional KL space. Subsequently, GP models are used to build the temporal system relationships. Thus, the nonlinear spatiotemporal dynamics can be reconstructed after the time/space synthesis. The advantage of the proposed model is that KL–GP provides the predictive distribution of the outputs and the estimate of the variance of its predicted outputs. The "active data" in the DPS region can be found for model improvement according to the predicted variances. Then the developed self-active KL–GP model is extended to include adaptation and on-line implementation in real time. Systematic design procedures are needed so that the DPS modeling problems can be solved because there are no guidelines to define the architecture needed for evolution in the traditional method. This is particularly good when reducing the computational demand of the DPS model. Simulation results of DPS are presented to demonstrate the effectiveness of the self-active KL–GP modeling method and the recursively selective KL–GP modeling method.

1. Introduction

Distributed parameter system (DPS) exists in many industry processes. DPS is distinguished by the fact that its states or variables are non-uniform in the space. The dynamics of this kind of system at particular spatial location and time are not only relevant to the previous system states, but also with other spatial locations. The coupling of time and space and infinite dimensionality makes them very difficult to analyze. However, for system prediction and controller design, the models of these spatiotemporal systems are always necessary.

The standard form of the spatiotemporal systems is the partial differential equations (PDEs) according to the mechanism of the physical, chemical or biological processes. This kind of model can predict the spatiotemporal outputs of the system accurately if the knowledge is good enough, while the infinite dimensional PDEs models which are used directly are not practical because of the practically limited sensors and computing power. In practice, finite dimensional models are usually used to approximate DPS. Li and Qi (2010) gave a review about spatiotemporal separating modeling methods of DPS. Several numerical solving methods have been developed to solve PDEs, including the finite elements method (Brenner and Scott, 2008), the finite difference method (Mitchell

* Corresponding authors. Tel.: +886 3 2654107; fax: +886 3 2654199.
E-mail addresses: jason@wavenet.cycu.edu.tw (J. Chen), leix@iipc.zju.edu.cn (L. Xie).

http://dx.doi.org/10.1016/j.ces.2014.10.032
0009-2509/

and Griffiths, 1980), the Galerkin method (Balas, 1983; Fletcher, 1984), the approximate inertial manifold method (Christofides and Daoutidis, 1997; Foias et al., 1988) etc. Some of the local solving approaches assume that the output at the specific locations is only determined by its neighboring locations and itself. The local models can be established using local data based on the lattice dynamical system (Guo and Billings, 2007; Mandelj et al., 2001). But how to determine the neighborhood is always difficult or uncertain. On the other hand, the global approaches utilize a set of global basis functions to express the DPS information. While different basis functions would impact the accuracy of modeling, such as Fourier series (Deng et al., 2005) and various kinds of polynomials (Boyd, 2001; Hussaini and Zang, 1987), these functions are not optimal and they do not vary with the real process.

Besides, establishing PDE models is relatively hard because the knowledge of the systems is usually insufficient. Thus, data-driven approaches are used in practice. Although traditional data-driven system identification methods for lumped parameter systems have been widely developed, modeling the spatiotemporal system is still in the infant stage of development. Known as proper orthogonal decomposition, Karhunen–Loève (KL) decomposition (Li and Qi, 2010) is a popular approach to finding the principal spatial structures. KL decomposition is a data-driven method which chooses a principal spatial structure of global basis functions. Moreover, these basis functions are orthogonal (Newman, 1996; Sirovich, 1987), which would make modeling more efficient. In this method, the spatial-temporal variables can be represented by infinite spatial basis functions $\{\phi_n(s)\}_{n=1}^{\infty}$ and the corresponding temporal coefficients $\{y_n(t)\}_{n=1}^{\infty}$, where $\{\phi_n(s)\}_{n=1}^{\infty}$ obtained by KL decomposition are orthogonal with each other (Newman, 1996; Sirovich, 1987). In fact, a finite number of spatial basis functions are able to provide a good approximation for most industrial processes. By utilizing this feature, only few low-dimensional temporal models are needed to be established. After the proper spatial basis functions are designed, temporal variables can be obtained by projecting the process data on these basis functions. To model the temporal dynamics, many traditional modeling methods have been proposed, such as the Hammerstein model (Qi and Li, 2009b), neural network model (Qi and Li, 2009a; Sahan et al., 1997), support vector machine (Qi et al., 2010). However, the above prediction models supply a scalar prediction only at any sampling point without any measure of the confidence in that prediction, because these models do not take into account the uncertainty of the model structure. Also, once the models are identified, one does not know how to measure the current model quality or how to select the valuable data to improve the trained model.

For the above problems, Gaussian process (GP) models are ideally suitable. They were developed in system modeling decades ago (Williams and Barber, 1998), though they actually have a long history in geostatistics (Cressie and Wikle, 2011; Finley et al., 2012) known as “kriging”. GP models have been increasingly used in nonlinear dynamic systems (Grancharova et al., 2008; Holsclaw et al., 2013), because of their capability of providing the uncertainties of the predictive outputs, as well as the relatively less number of optimizing parameters (Ažman and Kocijan, 2007, 2011). Recently, nonlinear curve regression using GP models has aroused more interest (Holsclaw et al., 2013) and been compared with other commonly used models. The GP models had been applied to nonlinear dynamic system modeling (Deisenroth et al., 2009; Gregorčič and Lightbody, 2008, 2009), biological processes (Gao et al., 2008), spectroscopic calibration (Chen and Wang, 2010), and particle dynamics (Hernandez and Grover, 2010). In practical process modeling, models are usually constructed offline before being applied to online processes. Whether the trained models are healthy and the predictions are reliable is crucial for process audit, advanced automatic control, and real-time optimization purposes. The performance improvement is influenced by the accuracy of the model. Unlike the conventional models, GP models are valuable because they provide the predictive variance of the predicted output. This means that if the predictive distribution is tightly packed, the confidence level of the model prediction is high; on the contrary, the prediction distribution over a wide range of values indicates that the model is highly uncertain.

In this work, the KL–GP model for DPS is proposed. KL decomposition is employed for separating the spatiotemporal DPSs to get the optimal basis functions and make the calculation easier. The GP models are established for the temporal dynamic model based on the temporal variables which are obtained by projecting process data onto the basis functions. The spatial-temporal output predictions along with their variance can be computed after the time/space synthesis of the spatial basis functions and the temporal predictions. Furthermore, for the real-time applications, based on the predictive variance at each time point and space locations, non-redundant data would be selected when the GP models are updated. In this research, the self-active (SA) data based KL–GP, called SA–KL–GP, is proposed. With the newly selected data, while KL–GP is retrained, it is by its very nature an off-line training (i.e., all the data are gathered and processed.) The calculations are repeated for increasingly more data, but the computation can become prohibitive because the computation of the kernel matrix decomposition is time-consuming and the dimensions of kernel matrix depend on the number of the collected data. Many recursive KL approaches had been proposed (Golub, 1973; Hall et al., 2002; Parlett, 1980), but none of them select the principal basis functions adaptively. In addition, the optimization of the GP model parameters is more time-consuming because of the supervised maximum likelihood approach. Although recursive GP training methods were studied (Ni et al., 2011, 2012), they are not exactly suitable for our structure. In this research, the extension of our recursive update of GP (Chan et al., 2013) to a recursive version of the KL–GP training model based on the active data, called recursively selective KL–GP (RS–KL–GP), is developed. It allows us to update the parameters of KL–GP when an active data pair is selected without using all the sampled data in the computation and without re-computing the kernel matrix and retraining GP. Thus, when applied to a given DPS, the objectives of the RS–KL–GP model formalism are to provide dynamic model architecture for on-line applications (more specifically, dynamic architecture means the model can self-adjust the number of spatial basis functions and update the GP parameters based on the selected active data in the changes of the process behavior).

The rest of this paper is organized as follows: In Section 2, a brief review of the preliminary knowledge of KL decomposition and GP models is given; then the KL–GP model is derived. The self-active data based KL–GP model is proposed in Section 3. Section 4 investigates how the adaptive technique is incorporated into the KL–GP model for on-line recursive identification. In each of these three sections, the same simulation problem is used to demonstrate the effectiveness of the proposed modeling method and to explain the features with the improved KL–GP model from Section 2 to Section 4. And finally conclusions are drawn in Section 5.

2. KL–GP model

Considering a nonlinear spatiotemporal process, the temporal input is $\mathbf{x}(t) \in R^D$ and the spatiotemporal output is $Y(s,t) \in R$, where D indicates the number of inputs of the spatiotemporal process, $s \in \Omega$ is the spatial variable, Ω is the spatial domain, and t is the temporal variable. The spatiotemporal variable $Y(s,t)$ is distributed