Engineering Applications of Artificial Intelligence 67 (2018) 211–225

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Cope with diverse data structures in multi-fidelity modeling: A Gaussian process method

Haitao Liu [a,*], Yew-Soon Ong [b,c], Jianfei Cai [b], Yi Wang [d]

[a] Rolls-Royce@NTU Corporate Lab, Nanyang Technological University, Singapore 637460, Singapore
[b] School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore
[c] Data Science and Artificial Intelligence Research Center, Nanyang Technological University, Singapore 639798, Singapore
[d] Applied Technology Group, Rolls-Royce Singapore, 6 Seletar Aerospace Rise, Singapore 797565, Singapore

ARTICLE INFO

Keywords:
Multi-fidelity modeling
Gaussian process regression
diverse data structures
knowledge transfer

ABSTRACT

Multi-fidelity modeling (MFM) frameworks, especially the Bayesian MFM, have gained popularity in simulation based modeling, uncertainty quantification and optimization, due to the potential for reducing computational budget. In the view of multi-output modeling, the MFM approximates the high-/low-fidelity outputs simultaneously by considering the output correlations, and particularly, it transfers knowledge from the inexpensive low-fidelity outputs that have many training points to enhance the modeling of the expensive high-fidelity output that has a few training points. This article presents a novel multi-fidelity Gaussian process for modeling with diverse data structures. The diverse data structures mainly refer to the diversity of high-fidelity sample distributions, i.e., the high-fidelity points may randomly fill the domain, or more challengingly, they may cluster in some subregions. The proposed multi-fidelity model is composed of a global trend term and a local residual term. Particularly, the flexible residual term extracts both the shared and output-specific residual information via a data-driven weight parameter. Numerical experiments on two synthetic examples, an aircraft example and a stochastic incompressible flow example reveal that this very promising Bayesian MFM approach is capable of effectively extracting the low-fidelity information for facilitating the modeling of the high-fidelity output using diverse data structures.

1. Introduction

In recent years, new advances in computers and computing science lead to the widespread use of computer simulation models, e.g., computational fluid dynamics (CFD) and finite element analysis (FEA), in engineering design and optimization. In simulation based engineering problems, surrogates are starting to play an important role, since they can approximate the expensive simulation model at some training points for alleviating computational burden. Gaussian process regression (Rasmussen and Williams, 2006), also known as Kriging (Lophaven et al., 2002), is a widely used surrogate model, since it can provide not only the prediction response but also the related prediction variance.

This article focuses on a multi-fidelity scenario where the simulator for the physics-based problem of interests can be run at multiple levels of fidelity. The high fidelity (HF) simulator yields the most accurate predictions but is most time-consuming; whereas the fast low fidelity (LF) simulators provide coarse predictions, which however include the main features of the problem and thus are useful for preliminary exploration. The LF simulators are usually simplified analysis models by using coarse finite element meshes, relaxed boundary or convergence conditions, etc. For example, it was reported by Benamara et al. (2016) that the HF simulation for a 1.5 stage booster has 5 million meshes and requires 2 h; but the LF simulation has only 0.7 million meshes and requires only 15 min. In practice, we cannot afford extensive HF simulations at many training points but many LF simulations are affordable. Suppose that the simulator has Q levels of fidelity, the multi-fidelity modeling (MFM), also known as variable-fidelity modeling or data fusion, attempts to utilize the knowledge from the correlated yet inexpensive $Q-1$ LF simulators to enhance the modeling of the expensive HF simulator.

Considering Q levels of fidelity as Q correlated outputs, the information fusion can be achieved in the multi-output modeling framework. The multi-output GP (MOGP), also known as multi-variate Kriging (Kleijnen and Mehdad, 2014), has been developed and investigated

* Corresponding author.
E-mail addresses: htliu@ntu.edu.sg (H. Liu), ASYSOng@ntu.edu.sg (Y.-S. Ong), ASJFCai@ntu.edu.sg (J. Cai), Yi.Wang4@Rolls-Royce.com (Y. Wang).

https://doi.org/10.1016/j.engappai.2017.10.008
Received 30 May 2017; Received in revised form 28 August 2017; Accepted 5 October 2017
0952-1976/

with a long history. The MOGP attempts to model multiple correlated outputs simultaneously by sharing the information across them, with the aim of outperforming individual modeling. The key in MOGP is to construct a valid multi-output covariance function to transfer useful information across outputs. A pioneer and well-known model developed in the field of geostatistics is called linear model of coregionalization (LMC) (Journel and Huijbregts, 1978). This model constructs valid covariance functions by a linear combination of several Gaussian processes. Thereafter, various MOGPs have been developed and extended in the context of LMC (Seeger et al., 2005; Bonilla et al., 2007; Hayashi et al., 2012; Osborne et al., 2012; Rakitsch et al., 2013; Dürichen et al., 2015; Hori et al., 2016). Another way to construct valid covariance functions is through process convolutions that convolve a base process, e.g., white Gaussian noise, with a smoothing kernel (Ver Hoef and Barry, 1998; Boyle and Frean, 2004; Álvarez and Lawrence, 2009, 2011).[1] The process convolutions can be regarded as a dynamic version of LMC (Álvarez and Lawrence, 2011; Álvarez et al., 2012).

In the multi-fidelity scenario, particularly, we attempt to use the inexpensive LF outputs to assist the modeling of the expensive HF output. Hence, compared to the typical MOGPs that share the information across the outputs, Kennedy and O'Hagan (2000) presented a Bayesian discrepancy-based MFM framework, which is an extension to the Co-Kriging model (Myers, 1982). In this framework, an auto-regressive model is proposed by expressing the HF output as the sum of the scaled LF outputs and an additive term that accounts for the discrepancy between the outputs, leading to not only the information sharing across outputs but also the asymmetric knowledge transfer from the LF outputs to the HF output. Later, Qian and Wu (2008) and Leen et al. (2012) provided an equivalent MFM framework in different views. Being a good multi-fidelity predictive model, Co-Kriging has been extended and improved, e.g., by reducing computational complexity (Le Gratiet and Garnier, 2014; Le Gratiet and Cannamela, 2015), using space-dependent scaling factor (Perdikaris et al., 2017), and incorporating gradient information (Han et al., 2013; Ulaganathan et al., 2015). Due to the remarkable performance, Co-Kriging has gained popularity in various fields, e.g., model inversion (Perdikaris and Karniadakis, 2016), uncertainty quantification (Perdikaris et al., 2015; Kennedy and O'Hagan, 2001), and multidisciplinary/robust/multi-objective optimization (Forrester et al., 2007; Keane, 2012; Han and Görtz, 2012; Kontogiannis et al., 2017). For more information about Co-Kriging and MFM, one can refer to the recent reviews and comparison studies (Fernández-Godino et al., 2016; Park et al., 2017; Toal, 2015).

In the context of Co-Kriging, it is usually assumed that we can control the sampling process such that the HF and LF training points spread over the entire domain evenly by for example the nested sampling strategy (Qian, 2009) and the nearest neighbor sampling strategy (Le Gratiet and Garnier, 2014). The space-filling nested data structure, though being beneficial for Co-Kriging, cannot always be available in practice. In realistic scenarios, we need to handle diverse data structures. The diverse data structures here mainly refer to the diversity of HF sample distributions, while the inexpensive LF outputs are assumed to have sufficient training points that cover the entire domain. For example, as shown in Fig. 1,[2] we have a set of uniformly distributed HF points, a set of randomly distributed HF points, and more challengingly, a set of partially distributed HF points clustered in a subregion. Besides, a practical example is that when using CFD solvers of different fidelities to simulate the flow around an aircraft, the inexpensive LF Euler simulation can be computed over the domain; while for saving computing time, the expensive HF Navier–Stokes simulation is only performed in flow regions with strong viscous effects. Hence, the diverse data structures, which contain different HF sample distributions, pose the demands of developing a particular multi-fidelity modeling approach that can effectively extract LF information to facilitate the HF modeling in different scenarios.

Therefore, this article presents a novel multi-fidelity GP model that is composed of a global trend term and a local residual term in order to tackle diverse data structures, e.g., full points that are available in the entire domain, or partial points that fill only some subregions. Particularly, in order to extract useful LF information effectively for facilitating the HF modeling, the local residual term contains a shared part and an output-specific part, the trade-off between which is dynamically determined by a data-driven weight parameter. The flexible model structure enables the approach to accomplish the multi-fidelity modeling well with diverse data structures.

The remainder of the article is organized as follows. Section 2 briefly introduces the single-output Gaussian process. Then, Section 3 presents the newly developed multi-fidelity Gaussian process in the MOGP framework. Thereafter, Section 4 comprehensively tests the new approach on two synthetic examples and two engineering examples with diverse characteristics and data structures. Finally, Section 5 offers some concluding remarks.

2. Single-output Gaussian process

Here we give a brief introduction to the single-output Gaussian process (SOGP). GP is a stochastic process wherein any finite subset of random variables follows a joint Gaussian distribution. As a non-parametric[3] model, the GP interprets the target function $f(\mathbf{x})$ where $\mathbf{x} \in R^d$ as a probability distribution in function space as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \tag{1}$$

which is completely defined by the mean function $m(\mathbf{x})$ that is usually taken as zero without loss of generality, and the covariance function $k(\mathbf{x}, \mathbf{x}')$. In practice, we usually use the squared exponential (SE) covariance function as

$$k_{SE}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}')^T P^{-1}(\mathbf{x}-\mathbf{x}')\right), \tag{2}$$

where the signal variance σ_f^2 represents an output scale amplitude; the ith element of the diagonal matrix $P \in R^{d\times d}$ is the characteristic length-scale l_i^2 that controls the width of the bell-shaped curve along the ith dimension. For other well-known covariance functions, e.g., the Matérn covariance function and the rational quadratic covariance function, one can refer to Rasmussen and Williams (2006).

Typically, in many realistic scenarios, instead of the latent function values themselves, we only have the observed response

$$y(\mathbf{x}) = f(\mathbf{x}) + \epsilon, \tag{3}$$

where the independent and identically distributed (*i.i.d.*) noise $\epsilon \sim \mathcal{N}(0, \sigma_s^2)$ accounts for the practical measurement errors, the modeling errors, the manufacturing tolerances, etc. It has been pointed out that we can gain benefits from the consideration of noise in GP for numerical stability (Ababou et al., 1994; Neal, 1997) and better statistical properties, e.g., prediction accuracy and coverage (Gramacy and Lee, 2012).

For the target function $f(\mathbf{x})$, suppose that we have a set of training points $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}^T$ in the design space $D \in [0,1]^d$, and their corresponding output observations $\mathbf{y} = \{y(\mathbf{x}_1), \ldots, y(\mathbf{x}_n)\}^T$. The joint prior distribution of the observed dataset $D = \{X, \mathbf{y}\}$ augmented with a test data $\{\mathbf{x}_*, f_*\}$ is as

$$\begin{bmatrix} \mathbf{y} \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} K(X,X)+\sigma_s^2 I & K(X,\mathbf{x}_*) \\ K(\mathbf{x}_*,X) & k(\mathbf{x}_*,\mathbf{x}_*) \end{bmatrix}\right), \tag{4}$$

[1] If the base process is a Gaussian process, then the convolved process is ensured to be a Gaussian process.

[2] The HF and LF functions in this 1D multi-fidelity example are expressed by Eqs. (45) and (46), respectively.

[3] "Non-parametric" means that the GP has no explicit parameters to control the functional form of the model. But it still has some *hyperparameters* that need to be inferred in the modeling process.