Computers and Chemical Engineering 106 (2017) 171–182

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

An adaptive sampling approach for Kriging metamodeling by maximizing expected prediction error

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

[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, 639798, Singapore

ARTICLE INFO

Article history:
Received 27 March 2017
Received in revised form 18 May 2017
Accepted 24 May 2017
Available online 8 June 2017

Keywords:
Adaptive sampling
Kriging metamodeling
Expected prediction error
Adaptive balance strategy

ABSTRACT

As a well-known approximation method, Kriging is widely used in process engineering design and optimization for saving computational budget. The Kriging model for a target function is fitted to a set of sample points, the responses of which are expensive to obtain in practice and the sample distribution of which has a great impact on the model prediction quality. Therefore, a main task in adaptive sampling for Kriging metamodeling is to gather informative points in order to build an accurate model with as few points as possible. To this end, we propose an adaptive sampling approach under the bias-variance decomposition framework. This novel sampling approach sequentially selects new points by maximizing an expected prediction error criterion that considers both the bias and variance information. Particularly, it presents an adaptive balance strategy to dynamically balance the local exploitation and global exploration via the error information from the previous iteration. Four benchmark cases and four engineering cases from low to high dimensions are used to assess the performance of the proposed approach. Numerical results reveal that this adaptive sampling approach is very promising for constructing accurate Kriging models for problems with diverse characteristics.

1. Introduction

As a statistical model, the Kriging metamodeling technique, also known as Gaussian process regression (GPR), has been extensively used in engineering design and optimization for relieving computational budget. For an expensive simulation-based function, Kriging can fit a mathematical model to a finite number of observed points as an approximation. The cheap-to-run Kriging model helps enhance the understanding of the target function by exploring the design space. Some variants of Kriging, e.g., co-Kriging (Kennedy and O'Hagan, 2000), blind-Kriging (Joseph et al., 2008), gradient-enhanced Kriging (Morris et al., 1993), and non-stationary Kriging (Xiong et al., 2007), have been developed for different purposes.

Given an expensive target function $f \in R^1$, the general Kriging metamodeling process consists of two parts: (1) generating a set of observed points by the design of experiments (DoE) techniques; and (2) fitting a Kriging model $\hat{f}$ to the observed data. It is found that the sample positions distinctly affect the prediction quality of Kriging. Considering the limited computational budget in practice, a key issue in Kriging metamodeling is how to gather informative points in order to build an accurate model with as few points as possible.

Suppose that we already have a set of initial points $\mathbf{X}_D$ in the domain $D \in R^n$, and maintain a large pool of candidate points $\mathbf{U} = D \setminus \mathbf{X}_D$. The main task of a sampling approach is to sequentially select informative points from $\mathbf{U}$ and evaluate their responses in order to efficiently refine the Kriging model. Recently, the adaptive sampling strategy, also known as active learning (Settles, 2010), has gained increasing attention. This sampling strategy sequentially selects new points based on the information of both the approximation model and the data itself from previous iterations. Recently, there have emerged various adaptive sampling strategies for global metamodeling (Mackman and Allen, 2010; Xu et al., 2014; Eason and Cremaschi, 2014; Garud et al., 2017; Wang and Ierapetritou, 2017). This article mainly focuses on the adaptive sampling developed under the Bayesian framework for Kriging metamodeling.

It is known that because of the Bayesian framework, the Kriging model provides not only the prediction response $\hat{f}(\boldsymbol{x})$ but also the prediction variance $s^2(\boldsymbol{x})$ (also known as mean square error, MSE) at an arbitrary point $\boldsymbol{x}$. Therefore, a straightforward adaptive sampling strategy for Kriging is to sequentially select a new point with

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

http://dx.doi.org/10.1016/j.compchemeng.2017.05.025

maximal prediction variance, denoted as maximum mean square error (MMSE) (Jin et al., 2002), in order to reduce the generalization error of $\hat{f}$. The MMSE sampling approach is widely used due to the ease of implementation and the cost-efficient process. Based on Shannon's entropy theory, Shewry and Wynn (1987) proposed the maximum entropy (ME) criterion to select new points by maximizing the determinant of the correlation matrix. Particularly, the one-by-one ME approach is equivalent to the MMSE approach (Jin et al., 2002). To further improve the sampling performance, Morris et al. (1993) combined the maximum entropy sampling approach with the known first derivatives. Besides, under the expected error reduction framework, Sacks et al. (1989) selected new points by maximizing the expected reduction in integrated mean square error (IMSE) over the entire design space, which is equivalent to the active learning-Cohn (ALC) criterion (Cohn, 1996; Christen et al., 2011). However, due to the integral operation, the computational complexity of IMSE is much higher than that of MSE. A recent work (Beck and Guillas, 2016) suggested using the mutual information to adaptively select new points by maximizing the expected information gain.

The points generated by the above variance-based sampling approaches are found to primarily fill the domain evenly and have a slight adaption to the variability along each coordinate direction. This is because that they consider only the Kriging variance information that follows a stationary assumption wherein the correlation function is identical over the entire domain. As a result, the prediction variance solely depends on the sample locations. For effectively improving the model accuracy within limited computational budget, rather than fill the domain evenly, an interesting idea is to determine the sample positions according to the characteristics of target function, e.g., sampling more points in regions with large prediction errors.

To this end, Lin et al. (2004) used the prediction errors estimated by additional validation points to adjust the correlation matrix so that the correlation function is no longer identical over the domain. That is, it owns the ability to adapt to the function characteristics, which thereafter helps the sampling approach identify regions with large prediction errors. The validation points, however, are usually unavailable in practice. In a similar spirit, Farhang-Mehr and Azarm (2005) employed the locations of local optima on current Kriging model to adjust the correlation matrix in order to identify irregular regions. This adaptive sampling criterion heavily depends on the quality of the Kriging model. A poor model may guide an erroneous sampling direction. Recently, Liu et al. (2016a) proposed an adaptive maximum entropy (AME) sampling approach by using the cross-validation errors to adjust the correlation function, and moreover, employing a user-defined search pattern to circularly conduct sampling from global to local. Busby (2009) and Busby et al. (2007) decomposed the domain into cells with the edges being of the order of correlation lengths along different directions. Then, they adopted the cross-validation criterion and ME criterion to identify "bad cells" for sampling. Besides, Lam (2008) modified the expected improvement criterion developed for global optimization (Jones et al., 1998) to obtain a good global model fit. The expected improvement for global fit (EIGF) approach selects informative points that have a large expected improvement over the nearest observed points.

As has been pointed out by Liu et al. (2016a) and Deschrijver et al. (2011), an effective adaptive sampling approach should contain three parts:

(1) *Local exploitation*. This part accounts for the adaption of the sampling process by guiding the sampling in regions with large prediction errors. The local exploitation term can be represented in various ways, e.g., the prediction errors at validation points (Lin et al., 2004), the cross-validation errors (Jin et al., 2002; Liu et al., 2016a; Busby et al., 2007), the locations of local optima (Farhang-Mehr and Azarm, 2005), the local response variation (Lam, 2008) and more others.
(2) *Global exploration*. This part ensures the purpose of global metamodeling, and avoids the missing of undetected interesting regions by employing, e.g., some distance-based criteria (Jin et al., 2002).
(3) *Trade-off between local & global*. Last but not least, this part balances the local exploitation and global exploration, and has a great impact on the sampling performance. Most of current adaptive sampling approaches for Kriging, however, usually use a fixed balance rule, which is non-beneficial for sampling performance. Liu et al. (2016a) introduced a user-defined search pattern to balance the local exploitation and global exploration, but it is still inflexible. For NURBs-based metamodeling, Turner et al. (2007) proposed using two pure global criteria and two pure local criteria to formulate a cooling schedule wherein the Bernstein basis functions are adopted to decide the dominant criterion in different sampling stages. Singh et al. (2013) illustrated three conceptual balance strategies and recommended the adaptive balance strategy.

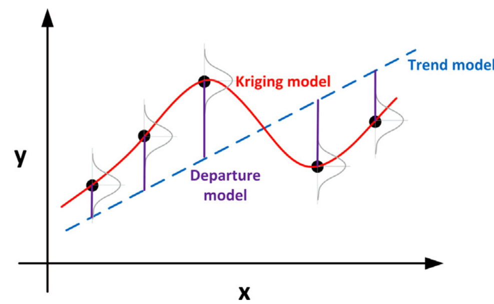

Fig. 1. Illustration of a Kriging model with linear regressions.

This article derives an adaptive sampling approach for Kriging metamodeling under the bias-variance decomposition framework. The proposed approach sequentially selects the most informative points through maximizing the expected prediction error criterion that considers both the bias and variance information. Besides, it presents a novel adaptive balance strategy to gain benefits from effective local exploitation, while not hurting the performance through dynamically balancing the local exploitation and global exploration via the error information from the previous iteration.

The remaining of the article is organized as follows. Section 2 gives a brief introduction of the Kriging model. Section 3 describes the proposed adaptive sampling approach via maximizing expected prediction error. Thereafter, four benchmark cases and four engineering cases from low to high dimensions are employed in Section 4 to assess the sampling performance. Finally, Section 5 offers some concluding remarks.

2. Kriging model

Kriging was first introduced in the field of geology to estimate the properties of sampled minerals given a set of sampled sites (Journel and Huijbregts, 1978). Thereafter, Sacks et al. (1989) applied Kriging in the context of design and analysis of computer experiments (DACE), the meaning of which now has been extended to refer to the suite of all metamodeling techniques (Viana et al., 2014).

In the Kriging framework, the main assumption is that the deterministic output $f(\boldsymbol{x})$ is regarded as the realization of a stochastic process $y_K(\boldsymbol{x})$. As shown in Fig. 1, Kriging is composed of a global polynomial model, called *trend model*, over the entire domain and