Knowledge-Based Systems 144 (2018) 102–121

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Remarks on multi-output Gaussian process regression

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, Singapore 639798, Singapore

A R T I C L E I N F O

Article history:
Received 18 August 2017
Revised 18 October 2017
Accepted 27 December 2017
Available online 3 January 2018

Keywords:
Multi-output Gaussian process
Symmetric/asymmetric MOGP
Multi-fidelity
Output correlation
Knowledge transfer

A B S T R A C T

Multi-output regression problems have extensively arisen in modern engineering community. This article investigates the state-of-the-art multi-output Gaussian processes (MOGPs) that can transfer the knowledge across related outputs in order to improve prediction quality. We classify existing MOGPs into two main categories as (1) symmetric MOGPs that improve the predictions for all the outputs, and (2) asymmetric MOGPs, particularly the multi-fidelity MOGPs, that focus on the improvement of high fidelity output via the useful information transferred from related low fidelity outputs. We review existing symmetric/asymmetric MOGPs and analyze their characteristics, e.g., the covariance functions (separable or non-separable), the modeling process (integrated or decomposed), the information transfer (bidirectional or unidirectional), and the hyperparameter inference (joint or separate). Besides, we assess the performance of ten representative MOGPs thoroughly on eight examples in symmetric/asymmetric scenarios by considering, e.g., different training data (heterotopic or isotopic), different training sizes (small, moderate and large), different output correlations (low or high), and different output sizes (up to four outputs). Based on the qualitative and quantitative analysis, we give some recommendations regarding the usage of MOGPs and highlight potential research directions.

1. Introduction

Computer simulators, e.g., computational fluid dynamics (CFD) and finite element analysis (FEA), have gained popularity in many scientific fields to simulate various physical problems. For computationally expensive simulators, we usually employ surrogates to approximate the input-output relationship in order to relieve computational budget [1–3]. As a statistical surrogate model that provides not only the predictions but also the relevant uncertainty, Gaussian process (GP) has been gaining widespread applications, e.g., small- or large-scale regression [4–6], dimensionality reduction [7], Bayesian optimization [8], uncertainty quantification [9] and time-series analysis [10].

Typical GPs are usually designed for single-output scenarios wherein the output is a scalar. However, the multi-output problems have arisen in various fields, e.g., environmental sensor networks [11], robot inverse dynamics [12], multivariate physiological time-series analysis [13], structural design [14], and aircraft design [15]. Suppose that we attempt to approximate T outputs $\{f_t\}_{1 \le t \le T}$, one intuitive idea is to use the single-output GP (SOGP) to approximate them individually using the associated training data $\mathcal{D}_t = \{X_t, \mathbf{y}_t\}$, see Fig. 1(a). Considering that the outputs are correlated in some way, modeling them individually may result in the loss of valuable information. Hence, an increasing diversity of engineering applications are embarking on the use of multi-output GP (MOGP), which is conceptually depicted in Fig. 1(b), for surrogate modeling.

The study of MOGP has a long history and is known as multivariate Kriging or Co-Kriging [16–19] in the geostatistic community; it also overlaps with the broad field of multi-task learning [20,21] and transfer learning [22,23] of the machine learning community. The MOGP handles problems with the basic assumption that the outputs are correlated in some way. Hence, a key issue in MOGP is to *exploit the output correlations such that the outputs can leverage information from one another* in order to provide more accurate predictions in comparison to modeling them individually.

Existing MOGPs can in general be classified into two categories: (1) *symmetric* MOGPs and (2) *asymmetric* MOGPs. Symmetric MOGPs use a *symmetric dependency structure* to capture the output correlations and approximate the T outputs simultaneously. Therefore, these MOGPs usually have an *integrated* modeling process, i.e., fusing all the information in an entire covariance matrix, which leads to *bidirectional* information transfer between the outputs. Typically, the symmetric MOGPs attempt to improve the predictions of all the outputs in symmetric scenarios, where

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

https://doi.org/10.1016/j.knosys.2017.12.034
0950-7051/

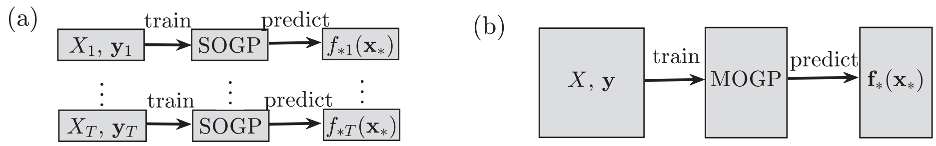

Fig. 1. Illustration of (a) the SOGP and (b) the MOGP.

the outputs are of equal importance and have roughly equivalent training information.

On the contrary, asymmetric MOGPs, which have an *asymmetric dependency structure* specifically designed for asymmetric scenarios, target to enhance the *primary* output predictions by transferring useful knowledge from other related *secondary* outputs.[1] The basic assumption is that the primary output has a few training points, but the secondary outputs, also denoted as source domains in transfer learning [24–26], usually have sufficient training points. Here, we particularly restrict ourselves to a *hierarchical* asymmetric scenario where the simulator for the physics-based problem of interests has multiple levels of fidelity. Regarding them as different outputs, the version with the highest fidelity is the primary output, which has been deem to give the most accurate predictions but is most time-consuming; whereas the simple and fast versions with declining fidelities provide coarse predictions, which however include the main features of the engineering problem and thus are useful for preliminary exploration. This kind of asymmetric multi-output modeling is often referred to as *multi-fidelity modeling* or *variable fidelity modeling* [27–29].

This article intends to (1) review and analyze the characteristics and differences of the state-of-the-art symmetric/asymmetric MOGPs, (2) investigate the potential of MOGPs over the typical SOGP on symmetric/asymmetric examples, and (3) give some recommendations regarding the usage of MOGPs.

The remainder of this article is organized as follows. Section 2 introduces the general single-/multi-output GP modeling framework. Thereafter, the existing symmetric and asymmetric MOGPs are reviewed and analyzed in Section 3 and Section 4, respectively. Section 5 further discusses the inference methods as well as the computational considerations to implement these MOGPs in practice. Subsequently, Section 6 investigates the performance and characteristics of symmetric MOGPs on four symmetric examples. Moreover, Section 7 studies the asymmetric/symmetric MOGPs on four asymmetric examples. Last, some concluding remarks are provided in Section 8.

2. Single-/multi-output Gaussian process modeling framework

In the multi-output scenario, assume that $X = \{\mathbf{x}_{t,i} | t = 1, \ldots, T; i = 1, \ldots, n_t\}$ and $\mathbf{y} = \{y_{t,i} = y_t(\mathbf{x}_{t,i}) | t = 1, \ldots, T; i = 1, \ldots, n_t\}$ are the collection of training points and associated observations for T outputs $\{f_t\}_{1 \le t \le T}$. Suppose that $N = \sum_{t=1}^{T} n_t$, the matrix $X \in R^{N \times d}$ has T blocks with the t-th block $X_t = \{\mathbf{x}_{t,1}, \ldots, \mathbf{x}_{t,n_t}\}^{\mathsf{T}}$ corresponding to the training set for output f_t; the vector $\mathbf{y} \in R^{N \times 1}$ also has T components with $\mathbf{y}_t = \{y_{t,1}, \ldots, y_{t,n_t}\}^{\mathsf{T}}$ corresponding to the observations of f_t at X_t.

Given the training data $\mathcal{D} = \{X, \mathbf{y}\}$ for T outputs, the task is to learn a MOGP model as

$$\text{MOGP} : \Omega_d \to \Omega_{f_1} \times \cdots \times \Omega_{f_T}$$

where Ω_d represents the d-dimensional input space, and Ω_{f_t} represents the output space for $f_t(\mathbf{x})$. In this article, we assume that all the T outputs share the same input space.[2] Besides, for the training sets, we consider two configurations below:

- *Heterotopic data*. It means the T outputs have different training sets, i.e., $X_1 \neq \cdots \neq X_T$. The heterotopic data often occurs in the scenario where the T output responses at a point $\mathbf{x}$ can be obtained by separate simulations.
- *Isotopic data*. It indicates the T outputs have the same training set, i.e., $X_1 = \cdots = X_T = \overline{X}$. The isotopic data often occurs in the scenario where the T output responses at a point $\mathbf{x}$ can be obtained simultaneously through a single simulation.

For the modeling of the outputs, this article considers two scenarios below:

- *Symmetric scenario*. In this scenario, the T outputs are of equal importance and have the same number of training points, i.e., $n_1 = \cdots = n_T = n$. This scenario attempts to improve the predictions of all the outputs and has been popularly studied in multi-output regression [30]. Both the heterotopic data and the isotopic data are available in this scenario, and the symmetric MOGPs can be used here.
- *Asymmetric scenario*. In this article, it particularly refers to the hierarchical multi-fidelity scenario where $n_1 > \cdots > n_T$. This scenario attempts to improve the predictions of the expensive high fidelity (HF) output f_T by transferring information from the inexpensive low fidelity (LF) outputs $\{f_t\}_{1 \le t \le T-1}$. Note that only the heterotopic training data occurs in this scenario, and both the symmetric and asymmetric MOGPs can be used here.

Throughout the section, we first introduce the typical single-output GP modeling framework, followed by the general multi-output GP modeling framework.

2.1. Single-output Gaussian process

Typical single-output GP (SOGP) attempts to approximate the target output $f(\mathbf{x})$ where $\mathbf{x} \in R^d$ by interpreting it 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 $m(\mathbf{x})$, which is usually taken as zero without loss of generality, and the covariance function $k(\mathbf{x}, \mathbf{x}')$. The well-known squared exponential (SE) covariance function [4], which is infinitely differentiable and smooth, is expressed as

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

where the signal variance σ_f^2 represents an output scale amplitude; the diagonal matrix $P \in R^{d \times d}$ contains the characteristic length scales $\{l_i^2\}_{1 \le i \le d}$ that represent the oscillation frequencies along different directions.

In many realistic scenarios, we only have the observation of the exact function value as

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

[1] Though symmetric MOGPs are available in asymmetric scenarios (see the illustration examples in [13]), they may be not so effective as the particularly designed asymmetric MOGPs.

[2] Some works have considered the scenario where the outputs have different input spaces, see the review paper [22].