Applied Mathematical Modelling 65 (2019) 586–596

Contents lists available at ScienceDirect

Applied Mathematical Modelling

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

Robust additive Gaussian process models using reference priors and cut-off-designs

Natalie Vollert[a,b,*], Michael Ortner[a], Jürgen Pilz[b]

[a] CTR Carinthian Tech Research AG, Europastraße 12, Villach 9524, Austria
[b] Department of Statistics, Alpen-Adria University of Klagenfurt, Universitätsstraße 65–67, Klagenfurt 9020, Austria

ARTICLE INFO

Article history:
Received 8 September 2017
Revised 29 June 2018
Accepted 31 July 2018
Available online 8 September 2018

Keywords:
Computer experiments
Robust estimation
Gaussian process models
Reference prior
Additive models

ABSTRACT

When powerful numerical tools like the finite element method encounter their limits for the evaluation of physical systems it is very common to use surrogate models as an approximation. There are many possible choices concerning the model approach, among which Gaussian process models are the most popular ones due to their clear statistical basis. A very desirable attribute of such surrogates is a high flexibility for making them applicable to a great class of underlying problems while obtaining interpretable results. To achieve this Gaussian processes are used as basis functions of an additive model in this work. Another important property of a surrogate is stability, which can be especially challenging when it comes to the estimation of the correlation parameters. To solve this we use a Bayesian approach where a reference prior is assigned to each component of the additive model assuring robust correlation matrices. Due to the additive structure of the model a simplified parameter estimation process is proposed that reduces the usually high-dimensional optimization problem to a few sub-routines of low dimension. Finally, we demonstrate this concept by modeling the magnetic field of a magnetic linear position detection system.

© 2018 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY license.
(http://creativecommons.org/licenses/by/4.0/)

1. Introduction

1.1. Background

The problem of finding an approximation for time-expensive computer models is especially common when dealing with multivariate optimization tasks in industrial applications. Such a surrogate model typically depends on all k input variables simultaneously, i.e. $y = f(x_1, \ldots, x_k)$. With rising dimension the necessary full space analysis to construct such a model becomes intractable due to the high number of required sample points. One way to avoid this curse of dimension is the application of simplified models with an additive structure.

* Corresponding author at: CTR Carinthian Tech Research AG, Europastraße 12, Villach, 9524, Austria.
E-mail addresses: natalie.vollert@ctr.at (N. Vollert), michael.ortner@ctr.at (M. Ortner), juergen.pilz@aau.at (J. Pilz).

https://doi.org/10.1016/j.apm.2018.07.050
0307-904X/© 2018 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license.
(http://creativecommons.org/licenses/by/4.0/)

The basis of additive models is the fact that each integrable function can be decomposed into summands with different dimensions of the form [1,2]

$$y := f(\mathbf{x}) = f_0 + \sum_{i=1}^{k} f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + \cdots + f_{12\ldots k}(x_1, \ldots, x_k), \tag{1}$$

where $\mathbf{x} = (x_1, \ldots, x_k)^T$ and y is the dependent output (response). This form is known as functional ANOVA representation [3] but also referred to as high dimensional model representation (HDMR) [4,5]. To avoid overfitting simpler models are often preferred to more complex ones. A famous class of models based only on first order terms $f_i(x_i)$; $i = 1, \ldots, k$; and an intercept f_0, thus less prone to overfitting, are the generalized additive models (GAM) [6], defined by

$$\mathbb{E}(Y|x_1, \ldots, x_k) = f_0 + \sum_{i=1}^{k} f_i(x_i), \tag{2}$$

where the $f_i(\cdot)$ are smooth basis functions and Y denotes the random response variable. Generally, regression models like GAM from classical statistics are not suitable for modeling computer experiments due to the lack of a well-defined random error [7]. Instead Gaussian process (GP) models, also referred to as Kriging models, are the approach of choice in this field. In its original form the correlation matrix R of such a process is defined by the product correlation rule, i.e. $\otimes_{i=1}^{k} R_{i|\theta_i,\nu_i}$ based on predefined one-dimensional correlation functions $r_i(x, x') := r(x, x'|\theta_i, \nu_i)$. While the roughness parameters ν_i are usually fixed and assumed constant for all dimensions, the range parameters θ_i need to be estimated. This estimation is again based on models of the form $y = f(x_1, \ldots, x_k)$ and hence struggles with the issues mentioned above. However, due to some basic properties of GPs it is also possible to build the correlation function within the framework of additive models, referred to as additive Gaussian process (AGP) modeling. This idea has originally been proposed by [8] and [9], using only first order terms in the sense of GAM and allowing interactions of all orders, respectively. An excellent overview of the advantages of the additive structure compared to usual GP models with product correlations, can be found in [10].

However, a general disadvantage of GP modeling that also remains in the additive structure is the instability of the likelihood when it comes to the estimation of the correlation parameters [11]. This happens if the parameters are either very large yielding a singular correlation matrix, but also if the matrix is near-diagonal, i.e. for estimates close to zero. A powerful approach proposed to stabilize these estimates is the use of non-informative reference priors [12]. These priors where first used in [13] for isotropic correlation functions as an improvement to Jeffrey's prior, which often yields improper posterior distributions. Subsequently, this approach was extended in [14] by also including a nugget effect, but still in an isotropic framework with a single range parameter. While most of this work was motivated by spatial data problems, the use of reference priors for modeling computer experiments based on grid designs with anisotropic product-correlations was investigated in [15]. Finally, the concept of robust estimation of correlation parameters in simulation studies by using reference priors for general latin hypercube designs was introduced in [12].

In this work we combine second order AGPs with reference priors to achieve both, the increasing flexibility of additive models and the stabilization of correlation parameter estimates. Furthermore, we test sampling designs with different attributes to determine optimal layouts for the special structure of these models. We show that the additive structure enables a simplified estimation scheme which is based on the relaxed maximum likelihood method [8]. In this case, the optimization for parameter estimation reduces to a few low-dimensional problems, which are easier to handle than a single high-dimensional one. After validating the proposed method for some well-known test functions we also demonstrate its predictive power for magnetic field modeling - a very important part of magnetic sensor system design.

1.2. Additive Gaussian process models

Let $Z(\mathbf{x}), \mathbf{x} \in \mathbb{R}^k$ be a Gaussian process, i.e. $Z(\mathbf{x}) \sim N(\mu(\mathbf{x}), K = \sigma^2 R)$ with mean function $\mu(\cdot)$, process variance σ^2 and correlation matrix R, generated by a correlation function $r(\mathbf{x}_i, \mathbf{x}_j), \mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^k$. AGP models are based on the fact that the direct sum, as well as the tensor product of covariance functions are again covariance functions: let Z_i, $i = 1, 2, 3$, be independent GPs with

$$\begin{aligned} Z_1 &\sim N(\mu_1, K_1) \\ Z_2 &\sim N(\mu_2, K_2) \\ Z_3 &\sim N(\mu_3, K_1 K_2), \end{aligned} \tag{3}$$

then it holds that $Z := Z_1 + Z_2 + Z_3$ is again a GP with

$$Z \sim N(\mu_1 + \mu_2 + \mu_3, K_1 + K_2 + K_1 K_2). \tag{4}$$

For the additive structure in Eq. (1) this means that each component function is independently modeled as a GP.

In a first step, each dimension gets assigned a univariate correlation function $r_i(x_i, x_i')$, $i = 1, \ldots, k$, where $\mathbf{x} = (x_1, \ldots, x_k)^T, \mathbf{x}' = (x_1', \ldots, x_k')^T \in \mathbb{R}^k$ are two different sample points and the process variances are assumed to be equal for all dimensions. The first order GPs f_i can then be simply defined as

$$f_i \sim N(\mu_i, \sigma^2 R_i) \tag{5}$$

Download English Version:

https://daneshyari.com/en/article/11032130

Download Persian Version:

https://daneshyari.com/article/11032130

Daneshyari.com