Technometrics, ISSN 0040-1706
Volume 59, number 3 (august 2017)

Additive Gaussian Process for Computer Models With Qualitative and Quantitative Factors

P. 283-292

X. Deng, C. Devon Lin, K.-W. Liu & R. K. Rowe

Abstract

Computer experiments with qualitative and quantitative factors occur frequently in various applications in science and engineering. Analysis of such experiments is not yet completely resolved. In this work, we propose an additive Gaussian process model for computer experiments with qualitative and quantitative factors. The proposed method considers an additive correlation structure for qualitative factors, and assumes that the correlation function for each qualitative factor and the correlation function of quantitative factors are multiplicative. It inherits the flexibility of unrestrictive correlation structure for qualitative factors by using the hypersphere decomposition, embracing more flexibility in modeling the complex systems of computer experiments. The merits of the proposed method are illustrated by several numerical examples and a real data application. Supplementary materials for this article are available online.

Bayesian Local Kriging

P. 293-304

Luc Pronzato & Maria-João Rendas

Abstract

We consider the problem of constructing metamodels for computationally expensive simulation codes; that is, we construct interpolators/predictors of functions values (responses) from a finite collection of evaluations (observations). We use Gaussian process (GP) modeling and kriging, and combine a Bayesian approach, based on a finite set GP models, with the use of localized covariances indexed by the point where the prediction is made. Our approach is not based on postulating a generative model for the unknown function, but by letting the covariance functions depend on the prediction site, it provides enough flexibility to accommodate arbitrary nonstationary observations. Contrary to kriging prediction with plug-in parameter estimates, the resulting Bayesian predictor is constructed explicitly, without requiring any numerical optimization, and locally adjusts the weights given to the different models according to the data variability in each neighborhood. The predictor inherits the smoothness properties of the covariance functions that are used and its superiority over plug-in kriging, sometimes also called empirical-best-linear-unbiased predictor, is illustrated on various examples, including the reconstruction of an oceanographic field over a large region from a small number of observations. Supplementary materials for this article are available online.

Selecting an Orthogonal or Nonorthogonal Two-Level Design for Screening

P. 305-318

Robert W. Mee, Eric D. Schoen & David J. Edwards

Abstract

This article presents a comparison of criteria used to characterize two-level designs for screening purposes. To articulate the relationships among criteria, we focus on 7-factor designs with 16–32 runs and 11-factor designs with 20–48 runs. Screening based on selected designs for each of the run sizes considered is studied with simulation using a forward selection procedure and the Dantzig selector. This article compares Bayesian D-optimal designs,

designs created algorithmically to optimize estimation capacity over various model spaces, and orthogonal designs by estimation-based criteria and simulation. In this way, we furnish both general insights regarding various design approaches, as well as a guide to make a choice among a few final candidate designs. Supplementary materials for this article are available online.

Effective Design-Based Model Selection for Definitive Screening Designs P. 319-329

Bradley Jones & Christopher J. Nachtsheim

Abstract

Since their introduction by Jones and Nachtsheim in 2011 Jones, B., and Nachtsheim, C. J. (2011), "A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects," *Journal of Quality Technology*, 43, 1–15.[Web of Science ®], [Google Scholar], definitive screening designs (DSDs) have seen application in fields as diverse as bio-manufacturing, green energy production, and laser etching. One barrier to their routine adoption for screening is due to the difficulties practitioners experience in model selection when both main effects and second-order effects are active. Jones and Nachtsheim showed that for six or more factors, DSDs project to designs in any three factors that can fit a full quadratic model. In addition, they showed that DSDs have high power for detecting all the main effects as well as one two-factor interaction or one quadratic effect as long as the true effects are much larger than the error standard deviation. However, simulation studies of model selection strategies applied to DSDs can disappoint by failing to identify the correct set of active second-order effects when there are more than a few such effects. Standard model selection strategies such as stepwise regression, all-subsets regression, and the Dantzig selector are general tools that do not make use of any structural information about the design. It seems reasonable that a modeling approach that makes use of the known structure of a designed experiment could perform better than more general purpose strategies. This article shows how to take advantage of the special structure of the DSD to obtain the most clear-cut analytical results possible.

An Exact Test of Fit for the Gaussian Linear Model Using Optimal Nonbipartite Matching P. 330-337

Samuel D. Pimentel, Dylan S. Small & Paul R. Rosenbaum

Abstract

Fisher tested the fit of Gaussian linear models using replicated observations. We refine this method by (1) constructing near-replicates using an optimal nonbipartite matching and (2) defining a distance that focuses on predictors important to the model's predictions. Near-replicates may not exist unless the predictor set is low-dimensional; the test addresses dimensionality by betting that model failures involve a subset of predictors important in the old fit. Despite using the old fit to pair observations, the test has exactly its stated level under the null hypothesis. Simulations show the test has reasonable power even when many spurious predictors are present.

Modeling Regression Quantile Process Using Monotone B-Splines P. 338-350

Yuan Yuan, Nan Chen & Shiyu Zhou

Abstract

Quantile regression as an alternative to conditional mean regression (i.e., least-square regression) is widely used in many areas. It can be used to study the covariate effects on the entire response distribution by fitting quantile regression models at multiple different quantiles or even fitting the entire regression quantile process. However, estimating the regression quantile process is inherently difficult because the induced conditional quantile function needs to be monotone at all covariate values. In this article, we proposed a regression quantile process estimation method based on monotone B-splines. The proposed method can easily ensure the validity of the regression quantile process and offers a concise framework for variable selection and adaptive complexity control. We thoroughly investigated the properties of the proposed procedure, both theoretically and numerically. We also used a case study on wind power generation to demonstrate its use and effectiveness in real problems. Supplementary materials for this article are available online.

Multiple Testing in Regression Models With Applications to Fault Diagnosis in the Big Data Era P. 351-360

Ching-Kang Ing, Tze Leung Lai, Milan Shen, KaWai Tsang & Shu-Hui Yu

Abstract

Motivated by applications to root-cause identification of faults in multistage manufacturing processes that involve a large number of tools or equipment at each stage, we consider multiple testing in regression models whose outputs represent the quality characteristics of a multistage manufacturing process. Because of the large number of input variables that correspond to the tools or equipments used, this falls in the framework of regression modeling in the modern era of big data. On the other hand, with quick fault detection and diagnosis followed by tool rectification, sparsity can be assumed in the regression model. We introduce a new approach to address the multiple testing problem and demonstrate its advantages over existing methods. We also illustrate its performance in an application to semiconductor wafer fabrication that motivated this development. Supplementary materials for this article are available online.

A Geometric Approach to Archetypal Analysis and Nonnegative Matrix Factorization P. 361-370

Anil Damle & Yuekai Sun

Abstract

Archetypal analysis and nonnegative matrix factorization (NMF) are staples in a statistician's toolbox for dimension reduction and exploratory data analysis. We describe a geometric approach to both NMF and archetypal analysis by interpreting both problems as finding extreme points of the data cloud. We also develop and analyze an efficient approach to finding extreme points in high dimensions. For modern massive datasets that are too large to fit on a single machine and must be stored in a distributed setting, our approach makes only a small number of passes over the data. In fact, it is possible to obtain the NMF or perform archetypal analysis with just two passes over the data.

Comparing the Reliability of Related Populations With the Probability of Agreement P. 371-380

Nathaniel T. Stevens & Christine M. Anderson-Cook

Abstract

Combining information from different populations to improve precision, simplify future predictions, or improve underlying understanding of relationships can be advantageous when considering the reliability of several related sets of systems. Using the probability of agreement to help quantify the similarities of populations can help to give a realistic assessment of whether the systems have reliability that are sufficiently similar for practical purposes to be treated as a homogeneous population. The new method is described and illustrated with an example involving two generations of a complex system, where the reliability is modeled using either a logistic or probit regression model. Note that supplementary materials including code, datasets, and added discussion are available online.

A Generalized Quasi-MMSE Controller for Run-to-Run Dynamic Models P. 381-390

Sheng-Tsaing Tseng & Pei-Yu Chen

Abstract

This study proposes a generalized quasi-minimum mean square error (qMMSE) controller for implementing a run-to-run process control where the process input–output relationship follows a general-order dynamical model with added noise. The expression of the process output, the long-term stability conditions and the optimal discount factor of this controller are derived analytically. Furthermore, we use the proposed second-order dynamical model to illustrate the effects of mis-identification of the process I-O model on the process total mean square error (TMSE). Via a comprehensive simulation study, the model demonstrates that the TMSE may inflate by more than 150% if a second-order dynamical model with moderately large carryover effects is wrongly identified as that of a first-order model. This means that the effects of mis-identification of the process I-O model on the process total mean square error (TMSE) is not negligible for implementing a dynamic run-to-run (RTR) process control. Supplementary materials for this article are available online.

Quantifying Nanoparticle Mixing State to Account for Both Location and Size Effects

P. 391-403

Ling Dong, Xiaodong Li, Dan Yu, Hui Zhang, Zhong Zhang, Yanjun Qian & Yu Ding

Abstract

Ripley's *K* function is commonly used to characterize the homogeneity of spatial point distribution. Not surprisingly, it becomes a favored tool in quantifying the nanoparticles mixing state in composite materials, a parameter that material scientists believe is of close relevance to certain properties of the nanoparticle-embedding material. Ripley's *K* function assumes that the spatial points are dimensionless. In reality, the nanoparticles, once mixed in a host material, form clusters or agglomerates of various sizes and shapes. Our analysis shows that using the original *K* function falls short of ranking or distinguishing the homogeneity of nanoparticle mixing. We therefore propose to revise the *K* function to account for both particle location and size effects. We apply the revised function to electron microscopy images of material samples and conduct analysis and comparison of nanoparticle mixing. The analysis shows that the revised function is a better index to quantify the mixing states.