15th Workshop on Quality Improvement Methods
at the Esplanade Hotel in Dortmund

Burgwall 3. D-44135 Dortmund
tel.: +49 231 58530 fax.: +49 231 5853270

technische universität
dortmund

SFB
823

ABSTRACTS

Online Selection of Surrogate Models for Constrained Black-Box Optimization
Samineh Bagheri, TH Köln University of Applied Sciences

Surrogate Model Based Optimization of the Substrate Feed Mixture of a Biogas Plant
Martin Zaefferer, TH Köln University of Applied Sciences

Optimal designs for survival data
Maria Konstantinou, Ruhr-Universität Bochum

Universally optimal designs under mixed interference models with and without block effects
Katarzyna Filipiak, Poznań University of Technology

A Physical-Statistical Model for Li-Ion Battery Prognostics
Chen Nan, National University of Singapore

Using the Six Sigma method to optimize Lithium-Ion-Battery production: A work in progress
Oliver Meyer, TU Dortmund University

Optimal design when outcome values may be missing
Stefanie Biedermann, University of Southampton

Optimal design of sampling survey for parameter estimation
Wei Zheng, Indiana University - Purdue University Indianapolis

Prediction intervals for failure processes
Christine Müller, TU Dortmund University

Martingale Methods with Applications in Reliability
Uwe Jensen, University of Hohenheim, Stuttgart

Bayesian Statistical Process Control in the Big Data era
Fabrizio Ruggeri, IMATI CNR

An EWMA-Type Sign Chart with Exact Run Length Properties
Philippe Castagliola, Université de Nantes & LS2N UMR CNRS 6004

Online Selection of Surrogate Models for Constrained Black-Box Optimization

Samineh Bagheri, TH Köln University of Applied Sciences

Real-world optimization problems are often subject to many constraints which are expensive to be evaluated in terms of cost and time. Surrogate- assisted optimization techniques aim to reduce evaluation costs by substituting the objective and the constraint functions with cheap surrogate models, e.g. radial basis functions.

Radial Basis Function (RBF) interpolations are cheap and can provide excellent fits with very few evaluated points, therefore, a common choice of surrogate for expensive optimization tasks. It has been shown that RBF with various basis functions (cubic, Gaussian, multiquadric,...) show contrary performance on modeling different classes of functions [1]. Selecting the correct basis function and its associated parameters is a challenging task and depends on many factors like initial design, population size and type of function to be modeled, which is usually unknown in black-box problems.

We propose an online model selection strategy that sorts different models according to their approximation error and selects in each iteration the best model for each function. The model selection strategy is embedded in the SACOBRA optimizer [2]. We show that the proposed online model selection strategy boosts the overall performance of the algorithm when applied to 24 well-known constrained optimization problems.

Keywords
Online model selection, Constrained optimization, Surrogate-assisted optimization

References

[1] Rocha, H. (2009): On the selection of the most adequate radial basis function, Applied Mathematical Modelling, 33(3):1573- 1583

[2] Bagheri, S., Konen, W., Emmerich, M., and Bäck, T. (2015): Solving the G-problems in less than 500 iterations: Improved efficient constrained optimization by surrogate modeling and adaptive parameter control. arXiv preprint arXiv:1512.09251.

Surrogate Model Based Optimization of Biogas Plant Feed Mixtures

Martin Zaefferer, TH Köln University of Applied Sciences

An essential task for operation and planning of biogas plants is the optimization of feed mixtures. Profitable operation of the plant requires the determination of the optimal amounts of maize, manure, grass silage, and other feedstocks. Due to the slowness of the underlying biochemical processes, optimization usually uses accurate simulation models. The simulation models themselves may be time-consuming to evaluate. Therefore, surrogate model-based approaches can be employed to optimize the feed mixtures efficiently.

To improve this model-based optimization approach, we present three ways to integrate prior knowledge into the optimization process. Firstly, a cheap, rough estimation is used to initialize the search. Secondly, the estimation is used to improve the surrogate model, via a multi-fidelity co-Kriging model. Thirdly, a two-layered modeling approach is used to avoid model deterioration due to potential discontinuities in the search space. Experimental results are presented, to outline the benefits as well as potential drawbacks of the described methods.

Optimal Designs for Survival Data

Maria Konstantinou, Ruhr-Universität Bochum

Survival data are observed in many industrial or biomedical 'time to event' experiments when the exact measurement of all the subjects is not observed. Finding efficient designs for such experiments can be problematic since the statistical models involved will usually be nonlinear, making the optimal choice of design parameter dependent. Analytical characterisations of locally D- and c-optimal designs are provided for a wide class of models, thus reducing the numerical effort for design search substantially. Designs which are robust to parameter misspecifications are also constructed when a range of parameter values is provided. Finally the issue of model robustness is adressed, since in practice the proposed parametric model will only hold approximately.

(Joint work with Stefanie Biedermann and Alan Kimber (University of Southampton))

Universally optimal designs under mixed interference models with and without block effects

Katarzyna Filipiak, Poznań University of Technology

Most of the papers about neighbor designs defined by Rees (1967) are devoted to the construction methods and only a few results to the statistical properties, such as e.g. efficiency or optimality, of proposed neighbor designs. A review of available literature, with special emphasize on the optimality of neighbor designs under various fixed effects interference models is given in Filipiak and Markiewicz (2017).

The aim of this paper is to verify, if the designs presented by Filipiak and Markiewicz (2017) as universally optimal under the fixed interference models are still universally optimal under the models with random interference effects. Moreover it is shown, that for a specified variance-covariance matrix of random interference effects, the universally optimal design under the mixed interference models with block effects is universally optimal over a wider class of designs.

In the paper the method presented by Filipiak and Markiewicz (2007) is extended and then applied to the mixed interference models without or with block effects.

References

Filipiak K., Markiewicz A. (2007): Optimal designs for a mixed interference model, Metrika, 65, 369-386.
Filipiak K., Markiewicz A (2017): Universally optimal designs under interference models with and without block effects, Comm Statist Theory Methods, 46, 1127-1143.
Rees DH (1967): Some designs of use in serology. Biometrics, 23, 779-791 (1967).

A Physical-Statistical Model for Li-Ion Battery Prognostics

Chen Nan, National University of Singapore

High intensity lithium-ion (Li-Ion) batteries have been widely used in consumer electronics, electric vehicles (EV), and critical marine and space systems. Their reliability and performance play an important role in the entire engineering system. As a result, developing prognostics and health management (PHM) approaches for Li-Ion batteries has received increasing attention in recent years.

This talk discusses a few model development for Li-Ion battery prognosis at different data and knowledge granularities. We use both simulation and real experiment data to demonstrate the features of different models. Through this research example, we want to highlight that statistical models, when fused with engineering knowledge, can lead to a more effective data analytics solution to practical problems.

Using the Six Sigma method to optimize Lithium-Ion-Battery production: A work in progress

Oliver Meyer, TU Dortmund University

Six Sigma offers a structured approach to optimize industrial processes based on statistical methods. As part of a BMBF (Federal Ministry of Education and Research) project to enhance the production of Lithium-Ion-Batteries we use Six Sigma to optimize the calendering process of coated foils, which serve as one of the main components of said type of batteries.

The process described here is implemented in the Research Production Line at the Center for Solar Energy and Hydrogen Research Baden-Württemberg in Ulm.

After optimizing the production process this production line is meant to serve as a blueprint for other state-of-the-art production lines of Lithium-Ion-Batteries in Germany.

Optimal design when outcome values may be missing

Stefanie Biedermann, University of Southampton

The presence of missing values complicates statistical analyses. In design of experiments, missing values are particularly problematic when constructing optimal designs, as it is not known which values are missing at the design stage. When data are missing at random (MAR) it is possible to incorporate this information into the optimality criterion that is used to find designs. However, when data are not missing at random (NMAR) such a framework can lead to inefficient designs.

We investigate and address the specific challenges that NMAR values present when finding optimal designs for linear regression models. We show that the optimality criteria will depend on model parameters that traditionally do not affect the design, such as regression coefficients and the residual variance. We also develop a framework that improves efficiency of designs over those found assuming values are missing at random.

(Joint work with Dr. Kim May Lee and Dr. Robin Mitra)

Optimal design of sampling survey for parameter estimation

Wei Zheng, Indiana University - Purdue University Indianapolis

For many tasks of data analysis, a large database of explanatory variables is readily available, however, the responses are missing and expensive to obtain. A natural remedy is to judiciously select a sample of the data, for which the responses are to be measured.

In this paper, we adopt the classical criteria in design of experiments to quantify the information of a given sample. Then, we provide a theoretical justification for approximating the optimal sample problem by a continuous problem, for which fast algorithms can be further developed with the guarantee of global convergence.

Our approach exhibits the following features:

1. The statistical efficiency of any candidate sample can be evaluated without knowing the exact optimal sample
2. It can be applied to a very wide class of statistical models
3. It can be integrated with a broad class of information criteria
4. It is scalable for big data.

Prediction intervals for failure processes

Christine Müller, TU Dortmund University

We consider the system of tension wires of a prestressed concrete beam under cyclic load where the time points of the breaks of the tension wires are known. As soon as a tension wire breaks, the load on the remaining tension wires increases. Hence we have a load sharing system where the increased load on the remaining tension wires provides an increased probability of a break of a remaining wire.

The increased probability of a break can be modeled by a birth process with a nonlinear intensity function. The nonlinear intensity function models also the influence of different external load to which the beams are exposed. Based on the former processes observed under the different external load levels, we want to predict the time point of a given number of failed tension wires for a new beam. The prediction should also work when the new beam is exposed to lower load than the former beams.

We present and compare different methods for getting prediction intervals for this problem. One method is based on the delta-method and three are based on confidence sets for the unknown parameters of the intensity function. One confidence set is obtained by data depth based on alternating signs. The comparison is done via cross validation for a real data set. At the end, we discuss the possibilities to include damage accumulation in the model.

Martingale Methods with Applications in Reliability

Uwe Jensen, Universität Hohenheim, Stuttgart

Survival Analysis comprises models which were developed to estimate lifetime or survival distributions in medicine or biology. Obviously these models can also be applied in other fields like, for example, reliability. Lifetime data are often censored, which means that because of a limited observation period lifetimes of individuals can only be observed within certain intervals. Such situations with reduced information due to censoring effects are taken into account by regression models in Survival Analysis.

The statistical inference of such point processes is carried out by modern tools of probability theory. Among these are martingale techniques which allow to model point and counting processes by means of intensities. Well known regression models which rely on martingale techniques are the Cox proportional hazards model and the additive model of O. Aalen. They are well suited if additional information should be entered into the model via covariates. In particular, in the standard Cox model the covariates are assumed to have constant influence. But some data sets exhibit deviations from this assumption.

Therefore, we examine a Cox-type regression model with change points in some of the covariates. Also a short overview on reliability prediction using regression models is given. This talk reviews some recent research with a personel perspective.

Bayesian Statistical Process Control in the Big Data era

Fabrizio Ruggeri, IMATI CNR

Quality control has been an activity which has long relied on a relatively limited number of observations and, in a Bayesian framework, on expert opinions on a very restricted set of parameters. Availability of large amount data and the increased complexity of industrial processes at hand (involving a remarkable number of parameters) provide new challenges in both monitoring many processes at the same time and using experts' knowledge to assess priors on the parameters of interest.

The talk will review the current literature in the field, discussing also directions for future research.

An EWMA-Type Sign Chart with Exact Run Length Properties

Philippe Castagliola, Université de Nantes & LS2N UMR CNRS 6004

In the literature, distribution-free control charts have been widely identified and investigated as easy and efficient means to monitor the location or the scale parameter of a process when the observations come from a non-normal or an unknown distribution.

In this work, a new Phase II EWMA-type sign chart for monitoring the location of an unknown continuous distribution is proposed and studied. The most valuable characteristics of this EWMA-type sign chart are that:

1. it only uses positive integer-valued weights to account for the past process history,
2. the plotted statistic is always an integer and
3. its run length properties can be exactly obtained without resorting to expensive simulations or unreliable approximations based on a continuous discretization of the discrete control chart limits.