Technometrics, ISSN 0040-1706
Volume 64, number 1 (2022)

Super Resolution for Multi-Sources Image Stream Data Using Smooth and Sparse Tensor Completion and Its Applications in Data Acquisition of Additive Manufacturing P. 2-17

Bo Shen, Rongxuan Wang, Andrew Chung Chee Law, Rakesh Kamath, Hahn Choo & Zhenyu (James) Kong

Abstract

Recent developments of advanced imaging systems spur their applications in many areas, ranging from satellite remote sensing for geographic information to thermal imaging analysis for manufacturing process monitoring and control. Due to different specifications of imaging systems, the resulting image stream data (videos) have different spatial and temporal resolutions. This proposed work is based on the image stream data captured by multiple imaging systems for the same object with different but complementary spatial and temporal resolutions. For example, one system has high spatial but low temporal resolutions while the other one has opposite resolutions. The goal of this article is to develop a new super resolution method that integrates these different types of image stream data to improve both spatial and temporal resolutions, which is critical to obtaining more insightful information for more effective quality control of targeted processes or systems. To fulfill this goal, a new tensor completion model is developed by considering both smooth and sparse features simultaneously and is thus termed smooth and sparse tensor completion (SSTC). The results of the extensive case studies illustrate the superiority of our method over the elaborately selected benchmark methods.

Individual Transition Label Noise Logistic Regression in Binary Classification for Incorrectly Labeled Data P. 18-29

Seokho Lee & Hyelim Jung

Abstract

We consider a binary classification problem in the case where some observations in the training data are incorrectly labeled. In the presence of such label noise, conventional classification fails to obtain a classifier to be generalized to a population. In this work, we investigate label noise logistic regression and explain how it works with noisy training data. We demonstrate that, when label transition probabilities are correctly provided, label noise logistic regression satisfies the Fisher consistency and enjoys the property of robustness. To accommodate various label noise mechanisms that occur in practice, we propose a flexible label noise model in a nonparametric way. We propose an efficient algorithm under the thresholding rule for individual parameter estimation. We demonstrate its performance under synthetic and real examples. We discuss the proposed flexible transition model is also useful for robust classification.

Robust and Efficient Parametric Spectral Density Estimation for High-Throughput Data P. 30-51

Martin Lysy, Feiyu Zhu, Bryan Yates & Aleksander Labuda

Abstract

Modern scientific instruments readily record various dynamical phenomena at high frequency and for extended durations. Spanning timescales across several orders of magnitude, such “high-throughput” (HTP) data are routinely analyzed with parametric models in the frequency domain. However, the large size of HTP datasets can render maximum likelihood estimation prohibitively expensive. Moreover, HTP recording devices are operated by extensive electronic circuitry, producing periodic noise to which parameter estimates are highly sensitive. This article proposes to address these issues with a two-stage approach. Preliminary parameter estimates are first obtained by a periodogram variance-stabilizing procedure, for which data compression greatly reduces computational costs with minimal impact to statistical efficiency. Next, a novel test with false discovery rate control eliminates most periodic outliers, to which the second-stage estimator becomes more robust. Extensive simulations and experimental results indicate that for a widely used model in HTP data analysis, a substantial reduction in mean squared error can be expected by applying our methodology.

High-Dimensional Cost-constrained Regression Via Nonconvex Optimization P. 52-64

Guan Yu, Haoda Fu & Yufeng Liu

Abstract

Budget constraints become an important consideration in modern predictive modeling due to the high cost of collecting certain predictors. This motivates us to develop cost-constrained predictive modeling methods. In this article, we study a new high-dimensional cost-constrained linear regression problem, that is, we aim to find the cost-constrained regression model with the smallest expected prediction error among all models satisfying a budget constraint. The nonconvex budget constraint makes this problem NP-hard. In order to estimate the regression coefficient vector of the cost-constrained regression model, we propose a new discrete first-order continuous optimization method. In particular, our method delivers a series of estimates of the regression coefficient vector by solving a sequence of 0-1 knapsack problems. Theoretically, we prove that the series of the estimates generated by our iterative algorithm converge to a first-order stationary point, which can be a globally optimal solution under some conditions. Furthermore, we study some extensions of our method that can be used for general statistical learning problems and problems with groups of variables. Numerical studies using simulated datasets and a real dataset from a diabetes study indicate that our proposed method can solve problems of fairly high dimensions with promising performance.

Computer Model Emulation with High-Dimensional Functional Output in Large-Scale Observing System Uncertainty Experiments P. 65-79

Pulong Ma, Anirban Mondal, Bledar A. Konomi, Jonathan Hobbs, Joon Jin Song & Emily L. Kang

Abstract

Observing system uncertainty experiments (OSUEs) have been recently proposed as a cost-effective way to perform probabilistic assessment of retrievals for NASA’s Orbiting Carbon Observatory-2 (OCO-2) mission. One important component in the OCO-2 retrieval algorithm is a full-physics forward model that describes the mathematical relationship between atmospheric variables such as carbon dioxide and radiances measured by the remote sensing instrument. This complex forward model is computationally expensive but large-scale OSUEs require evaluation of this model numerous times, which makes it infeasible for comprehensive experiments. To tackle this issue, we develop a statistical emulator to facilitate large-scale OSUEs in the OCO-2 mission. Within each distinct spectral band, the emulator represents radiance output at irregular wavelengths as a linear combination of basis functions and random coefficients. These random coefficients are then modeled with nearest-neighbor Gaussian processes with built-in input dimension reduction via active subspace. The proposed emulator reduces dimensionality in both input space and output space, so that fast computation is achieved within a fully Bayesian inference framework. Validation experiments demonstrate that this emulator outperforms other competing statistical methods and a reduced order model that approximates the full-physics forward model.

An Information Geometry Approach to Robustness Analysis for the Uncertainty Quantification of Computer Codes P. 80-91

Clement Gauchy, Jerome Stenger, Roman Sueur & Bertrand Iooss

Abstract

Robustness analysis is an emerging field in the uncertainty quantification domain. It involves analyzing the response of a computer model—which has inputs whose exact values are unknown—to the perturbation of one or several of its input distributions. Practical robustness analysis methods therefore require a coherent methodology for perturbing distributions; we present here one such rigorous method, based on the Fisher distance on manifolds of probability distributions. Further, we provide a numerical method to calculate perturbed densities in practice which comes from Lagrangian mechanics and involves solving a system of ordinary differential equations. The method introduced for perturbations is then used to compute quantile-related robustness indices. We illustrate these "perturbed-law based" indices on several numerical models. We also apply our methods to an industrial setting: the simulation of a loss of coolant accident in a nuclear reactor, where several dozen of the model's physical parameters are not known exactly, and where limited knowledge on their distributions is available.

Gaussian Process-Aided Function Comparison Using Noisy Scattered Data P. 92-102

Abhinav Prakash, Rui Tuo & Yu Ding

Abstract

This work proposes a nonparametric method to compare the underlying mean functions given two noisy datasets. The motivation for the work stems from an application of comparing wind turbine power curves. Comparing wind turbine data presents new problems, namely the need to identify the regions of difference in the input space and to quantify the extent of difference that is statistically significant. Our proposed method, referred to as funGP, estimates the underlying functions for different data samples using Gaussian process models. We build a confidence band using the probability law of the estimated function differences under the null hypothesis. Then, the confidence band is used for the hypothesis test as well as for identifying the regions of difference. This identification of difference regions is a distinct feature, as existing methods tend to conduct an overall hypothesis test stating whether two functions are different. Understanding the difference regions can lead to further practical insights and help devise better control and maintenance strategies for wind turbines. The merit of funGP is demonstrated by using three simulation studies and four real wind turbine datasets.

Accounting for Location Measurement Error in Imaging Data With Application to Atomic Resolution Images of Crystalline Materials P. 103-113

Matthew J. Miller, Matthew J. Cabral, Elizabeth C. Dickey, James M. LeBeau & Brian J. Reich

Abstract

Scientists use imaging to identify objects of interest and infer properties of these objects. The locations of these objects are often measured with error, which when ignored leads to biased parameter estimates and inflated variance. Current measurement error methods require an estimate or knowledge of the measurement error variance to correct these estimates, which may not be available. Instead, we create a spatial Bayesian hierarchical model that treats the locations as parameters, using the image itself to incorporate positional uncertainty. We lower the computational burden by approximating the likelihood using a noncontiguous block design around the object locations. We use this model to quantify the relationship between the intensity and displacement of hundreds of atom columns in crystal structures directly imaged via scanning transmission electron microscopy (STEM). Atomic displacements are related to important phenomena such as piezoelectricity, a property useful for engineering applications like ultrasound. Quantifying the sign and magnitude of this relationship will help materials scientists more precisely design materials with improved piezoelectricity. A simulation study confirms our method corrects bias in the estimate of the parameter of interest and drastically improves coverage in high noise scenarios compared to non-measurement error models.

Industrial Forecasting with Exponentially Smoothed Recurrent Neural Networks P. 114-124

Matthew Dixon

Abstract

Time series modeling has entered an era of unprecedented growth in the size and complexity of data which require new modeling approaches. While many new general purpose machine learning approaches have emerged, they remain poorly understand and irreconcilable with more traditional statistical modeling approaches. We present a general class of exponential smoothed recurrent neural networks (RNNs) which are well suited to modeling nonstationary dynamical systems arising in industrial applications. In particular, we analyze their capacity to characterize the nonlinear partial autocorrelation structure of time series and directly capture dynamic effects such as seasonality and trends. Application of exponentially smoothed RNNs to forecasting electricity load, weather data, and stock prices highlight the efficacy of exponential smoothing of the hidden state for multistep time series forecasting. The results also suggest that popular, but more complicated neural network architectures originally designed for speech processing are likely over-engineered for industrial forecasting and light-weight exponentially smoothed architectures, trained in a fraction of the time, capture the salient features while being superior and more robust than simple RNNs and autoregressive models. Additionally, uncertainty quantification of Bayesian exponential smoothed RNNs is shown to provide improved coverage.

Prediction of Future Failures for Heterogeneous Reliability Field Data P. 125-138

Colin Lewis-Beck, Qinglong Tian & William Q. Meeker

Abstract

This article introduces methods for constructing prediction bounds or intervals for the number of future failures from heterogeneous reliability field data. We focus on within-sample prediction where early data from a failure-time process is used to predict future failures from the same process. Early data from high-reliability products, however, often have limited information due to some combination of small sample sizes, censoring, and truncation. In such cases, we use a Bayesian hierarchical model to model jointly multiple lifetime distributions arising from different subpopulations of similar products. By borrowing information across subpopulations, our method enables stable estimation and the computation of corresponding prediction intervals, even in cases where there are few observed failures. Three applications are provided to illustrate this methodology, and a simulation study is used to validate the coverage performance of the prediction intervals.