Technometrics, ISSN 0040-1706
Volume 63, number 3 (2021)

Functional Regression Control Chart

P. 281-294

Fabio Centofanti, Antonio Lepore, Alessandra Menafoglio, Biagio Palumbo & Simone Vantini

Abstract

The modern development of data acquisition technologies in many industrial processes is facilitating the collection of quality characteristics that are apt to be modeled as functions, which are usually referred to as profiles. At the same time, measurements of concurrent variables, which are related to the quality characteristic profiles, are often available in a functional form as well, and usually referred to as covariates. To adjust the monitoring of the quality characteristic profiles by the effect of this additional information, a new functional control chart is elaborated on the residuals obtained from a function-on-function linear regression of the quality characteristic profile on the functional covariates. By means of a Monte Carlo simulation study, the proposed control chart is compared with other control charts already appeared in the literature and some remarks are given on its use in presence of covariate mean shifts. Furthermore, a real-case study in the shipping industry is presented with the purpose of monitoring ship fuel consumption and thus, CO2 emissions from a Ro-Pax ship, with particular regard to detecting their reduction after a specific energy efficiency initiative.

An Intrinsic Geometrical Approach for Statistical Process Control of Surface and Manifold Data

P. 295-312

Xueqi Zhao & Enrique del Castillo

Abstract

We present a new method for statistical process control (SPC) of a discrete part manufacturing system based on intrinsic geometrical properties of the parts, estimated from three-dimensional sensor data. An intrinsic method has the computational advantage of avoiding the difficult part registration problem, necessary in previous SPC approaches of three-dimensional geometrical data, but inadequate if noncontact sensors are used. The approach estimates the spectrum of the Laplace–Beltrami (LB) operator of the scanned parts and uses a multivariate nonparametric control chart for online process control. Our proposal brings SPC closer to computer vision and computer graphics methods aimed to detect large differences in shape (but not in size). However, the SPC problem differs in that small changes in either shape or size of the parts need to be detected, keeping a controllable false alarm rate and without completely filtering noise. An online or “Phase II” method and a scheme for starting up in the absence of prior data (“Phase I”) are presented. Comparison with earlier approaches that require registration shows the LB spectrum method to be more sensitive to rapidly detect small changes in shape and size, including the practical case when the sequence of part datasets is in the form of large, unequal size meshes. A post-alarm diagnostic method to investigate the location of defects on the surface of a part is also presented. While we focus in this article on surface (triangulation) data, the methods can also be applied to point cloud and voxel metrology data.

Adaptive Process Monitoring Using Covariate Information P. 313–328

Kai Yang & Peihua Qiu

Abstract
Statistical process control (SPC) charts provide a powerful tool for monitoring production lines in manufacturing industries. They are also used widely in other applications, such as sequential monitoring of internet traffic flows, disease incidences, health care systems, and more. In practice, quality/performance variables are often affected in a complex way by many covariates, such as material, labor, weather conditions, social/economic conditions, and so forth. Among all these covariates, some could be observed, some might be difficult to observe, and the others might even be difficult for us to notice their existence. Intuitively, an SPC chart could be improved by using helpful information in covariates. However, because of the complex relationship between the quality/performance variables and the covariates, shifts in the quality/performance variables could be due to certain covariates whose data cannot be collected. On the other hand, shifts in some observable covariates may not necessarily cause shifts in the quality/performance variables. Thus, it is challenging to properly use covariate information for process monitoring in a general setting. This article suggests a method to handle this problem. An effective exponentially weighted moving average chart is developed, in which its weighting parameter is chosen large if the related covariates included in the collected data tend to have a shift and small otherwise. Because the covariate information is used in the weighting parameter only, the chart is designed solely for detecting shifts in the quality/performance variables, but it can react to a future shift in the quality/performance variables quickly because the helpful covariate information has been used in its observation weighting mechanism. Extensive numerical studies show that this method is effective in many different cases.

Gaussian Process Assisted Active Learning of Physical Laws P. 329-342

Jiuhai Chen, Lulu Kang & Guang Lin

Abstract
In many areas of science and engineering, discovering the governing differential equations from the noisy experimental data is an essential challenge. It is also a critical step in understanding the physical phenomena and prediction of the future behaviors of the systems. However, in many cases, it is expensive or time-consuming to collect experimental data. This article provides an active learning approach to estimate the unknown differential equations accurately with reduced experimental data size. We propose an adaptive design criterion combining the D-optimality and the maximin space-filling criterion. In contrast to active learning for other regression models, the D-optimality here requires the unknown solution of the differential equations and derivatives of the solution. We estimate the Gaussian process (GP) regression models from the available experimental data and use them as the surrogates of these unknown solution functions. The derivatives of the estimated GP models are derived and used to substitute the derivatives of the solution. Variable selection-based regression methods are used to learn the differential equations from the experimental data. Through multiple case studies, we demonstrate the proposed approach outperforms the D-optimality and the maximin space-filling design alone in terms of model accuracy and data economy.

Order-Constrained ROC Regression With Application to Facial Recognition P. 343-353

Xiaochen Zhu, Martin Slawski, P. Jonathon Phillips & Liansheng Larry Tang

Abstract

The receiver operating characteristic (ROC) curve is widely used to assess discriminative accuracy of two groups based on a continuous score. In a variety of applications, the distributions of such scores across the two groups exhibit a stochastic ordering. Specific examples include calibrated biomarkers in medical diagnostics or the output of matching algorithms in biometric recognition. Incorporating stochastic ordering as an additional constraint into estimation can improve statistical efficiency. In this article, we consider modeling of ROC curves using both the order constraint and covariates associated with each score given that the latter (e.g., demographic characteristics of the underlying subjects) often have a substantial impact on discriminative accuracy. The proposed method is based on the indirect ROC regression approach using a location-scale model, and quadratic optimization is used to implement the order constraint. The statistical properties of the proposed order-constrained least squares estimator are studied. Based on the theoretical results developed herein, we deduce that the proposed estimator can achieve substantial reductions in mean squared error relative to its unconstrained counterpart. Simulation studies corroborate the superior performance of the proposed approach. Its practical usefulness is demonstrated in an application to face recognition data from the "Good, Bad, and Ugly" face challenge, a domain in which accounting for covariates has hardly been studied.

Lu Lu, Bingxing Wang, Yili Hong & Zhisheng Ye

Abstract

Degradation data have been broadly used for assessing product and system reliability. Most existing work focuses on modeling and analysis of degradation data with a single characteristic. In some degradation tests, interest lies in measuring multiple characteristics of the product degradation process to understand different aspects of the reliability performance, resulting in degradation data with multiple characteristics. The literature on modeling such data is scarce. Motivated by the photodegradation process of polymeric material, we propose a multivariate general path model for analyzing degradation data with multiple degradation characteristics (DCs). The model incorporates covariates for modeling the nonlinear degradation path. It also includes random effects that are correlated among the multiple DCs to capture the unit-to-unit variation in the individual degradation paths and to model the interdependence among the multivariate measurements. An expectation-maximization algorithm combined with the Markov chain Monte Carlo simulation is developed for estimating the model parameters and predicting system reliability with quantified uncertainty. The performance of the developed method is evaluated and compared with existing methods through a simulation study. The implementation of the method is illustrated through two examples with different forms of reliability functions. The main motivating example analyzes the coating degradation data with a closed-form reliability function, while the second example on analyzing the Device-B data demonstrates a more general simulation approach for dealing with analytically intractable reliability functions.

Chen Zhang, Hao Yan, Seungho Lee & Jianjun Shi

Abstract

Multivariate functional data from a complex system are naturally high-dimensional and have a complex cross-correlation structure. The complexity of data structure can be observed as that (1) some functions are strongly correlated with similar features, while some others may have almost no cross-correlations with quite diverse features; and (2) the cross-correlation structure may also change over time due to the system evolution. With this regard, this article presents a dynamic subspace learning method for multivariate functional data modeling. In particular, we consider that different functions come from different subspaces, and only functions of the same subspace have cross-correlations with each other. The subspaces can be automatically formulated and learned by reformatting the problem as a sparse regression. By allowing but regularizing the regression change over time, we can describe the cross-correlation dynamics. The model can be efficiently estimated by the fast iterative shrinkage-thresholding algorithm, and the features of each subspace can be extracted using the smooth multi-channel functional principal component analysis. Some theoretical properties of the model are presented. Numerical studies, together with case studies, demonstrate the efficiency and applicability of the proposed methodology.

Jialei Chen, Simon Mak, V. Roshan Joseph & Chuck Zhang

Abstract

Three-dimensional printed medical prototypes, which use synthetic metamaterials to mimic biological tissue, are becoming increasingly important in urgent surgical applications. However, the mimicking of tissue mechanical properties via three-dimensional printed metamaterial can be difficult and time-consuming, due to the functional nature of both inputs (metamaterial structure) and outputs (mechanical response curve). To deal with this, we propose a novel function-on-function kriging model for efficient emulation and tissue-mimicking optimization. For functional inputs, a key novelty of our model is the spectral-distance (SpeD) correlation function, which captures important spectral differences between two functional inputs. Dependencies for functional outputs are then modeled via a co-kriging framework. We further adopt shrinkage priors on both the input spectra and the output co-kriging covariance matrix, which allows the emulator to learn and incorporate important physics (e.g., dominant input frequencies, output curve properties). Finally, we demonstrate the effectiveness of the proposed SpeD emulator in a real-world study on mimicking human aortic tissue, and show that it can provide quicker and more accurate tissue-mimicking performance compared to existing methods in the medical literature.

Harjit Hullait, David S. Leslie, Nicos G. Pavlidis & Steve King

Abstract

Functional linear regression is a widely used approach to model functional responses with respect to functional inputs. However, classical functional linear regression models can be severely affected by outliers. We therefore introduce a Fisher-consistent robust functional linear regression model that is able to effectively fit data in the presence of outliers. The model is built using robust functional principal component and least squares regression estimators. The performance of the functional linear regression model depends on the number of principal components used. We therefore introduce a consistent robust model selection procedure to choose the number of principal components. Our robust functional linear regression model can be used alongside an outlier detection procedure to effectively identify abnormal functional responses. A simulation study shows our method is able to effectively capture the regression behavior in the presence of outliers, and is able to find the outliers with high accuracy. We demonstrate the usefulness of our method on jet engine sensor data. We identify outliers that would not be found if the functional responses were modeled independently of the functional input, or using nonrobust methods.

Jaesung Lee, Shiyu Zhou & Junhong Chen

Abstract

Two-dimensional layered materials/flakes, also known as crystalline atom-thick layer nanosheets, have recently been receiving great attention in electronics fabrication due to their unique and intriguing properties. The k-layer coverage area (i.e., the area covered by k number of overlapping layers) of the printed flake pattern significantly impacts on the properties of the printed electronics. In this work, we constructed a statistical model to describe the k-layer coverage of randomly distributed two-dimensional materials. A series of results are obtained to provide not only the expectation but also the variance of the coverage area. The boundary effects on the random flakes coverage are also studied. In addition, an approximated statistical testing approach is also developed in this work to detect abnormal coverage patterns. The case studies based on simulated data and real flakes images obtained from the inkjet printing process demonstrate the accuracy and effectiveness of the proposed model and analysis methods.

Technometrics, ISSN 0040-1706
Volume 63, number 3 (2021)