Mean Shift for Clustering Functional Data: A Scalable Algorithm and Convergence Analysis

Ting-Li Chen[1], **Toshinari Morimoto[2]**, Su-Yun Huang[1], Ruey S. Tsay[3]

[1]*Institute of Statistical Science, Academia Sinica*

[2]*Department of Mathematics, National Taiwan University*

[3]*Booth School of Business, University of Chicago*

ABSTRACT

This paper extends the mean shift algorithm from vector-valued data to functional data, enabling effective clustering in infinite-dimensional settings. To address the computational challenges posed by large-scale datasets, we introduce a fast stochastic variant that significantly reduces computational complexity. We provide a rigorous analysis of convergence for the full functional mean shift procedure, establishing theoretical guarantees for its behavior. For the stochastic variant, we provide some partial justification for its use by showing that it approximates the full algorithm well when the subset size is sufficiently large. The proposed method is validated both through simulation studies and through real-data analysis, including hourly Taiwan PM2.5 measurements and Argo oceanographic profiles. Our key contributions include:

(1) a novel extension of the mean shift algorithm to functional data for clustering without the need to specify the number of clusters

(2) convergence analysis of the full functional mean shift algorithm in Hilbert space

(3) a scalable stochastic variant based on random partitioning, with partial theoretical justification

(4) real-data applications demonstrating the method's scalability and practical usefulness
To ensure consistency in the online technical program, this abstract template must be used for submissions.

Keywords: mean shift clustering, functional data analysis, big data, convergence analysis, randomized algorithm

Personalized Functional Principal Component Analysis with Applications

Ruey S. Tsay[1,2], Ming-Chung Chang[3], Yen-Shiu Chin[3]

[1]*Booth School of Business, University of Chicago, Chicago, IL, USA*

[2]*Institute of Statistics and Data Science, National Tsing-Hua University, Hsinchu, Taiwan*

[3]*Institute of Statistical Science, Academia Sinica, Taipei, Taiwan*

ABSTRACT

Large spatio-temporal functional data are widely available nowadays. They are important in investigating global warming and climate changes. Dimension reduction becomes essential in studying those large functional data and functional principal component analysis is one of the commonly used dimensional reduction methods. On the other hand, similarly to other big data, large spatio-temporal functional data often exhibit certain commonality and some specific local features. The conventional functional principal component analysis becomes inadequate to handle such heterogeneous functional data. In this talk, we generalize personalized PCA to personalized functional PCA. We address both the computational and theoretical issues. For applications, we apply the proposed personalized functional PCA to the PM2.5 measurements in Taiwan and ARGO data.

Keywords: Functional principal component analysis, Functional data, Climate change, Dimension reduction

Nonstationary Gaussian Scale Mixture Models for Spatial Functional Extremes

Yen-Shiu Chin[1], Hsin-Cheng Huang[1], Nan-Jung Hsu[2]

[1]*Institute of Statistical Science, Academia Sinica*

[2]*Institute of Statistics and Data Science, National Tsing Hua University*

ABSTRACT

Extreme events are rare but highly impactful. For example, the prolonged heavy rainfall lasting five hours on June 20, 2021, in southern Taiwan caused agricultural losses exceeding NTD 75 million. Spatial modeling of such extreme environmental events viewed as realizations of functional processes defined over continuous spatial domains is essential for risk assessment in climatology, hydrology, and related fields.

Gaussian scale mixture (GSM) models, constructed as the product of a standard Gaussian process and a positive random scale, provide great flexibility for modeling spatial extremes. An important property of GSM models is that the tail behavior of the random scale determines their extremal dependence structure. Heavy-tailed scales, such as Pareto- or inverse-Gamma-type behavior, lead to asymptotic dependence through extreme shocks in the mixture and simultaneous occurrences of extreme events, whereas lighter-tailed scales, such as Weibull- or Rayleigh-type distributions, yield asymptotic independence. Despite this flexibility, existing GSM models are generally restricted to stationary or isotropic forms, limiting their applicability to complex environmental phenomena.

In this talk, we propose a new class of nonstationary Gaussian scale mixture models using a basis-function approach. Model parameters are estimated via maximum likelihood, and the computation is implemented efficiently through an expectation-maximization algorithm. We demonstrate the advantages of the proposed method through simulation studies and an environmental application to Irish temperature extremes.
Our findings show that allowing nonstationarity leads to more realistic spatial representations of environmental extremes and provides a modern functional data perspective for modeling spatial risk in environmental sciences.

Keywords: extremal dependence structure, Gaussian scale mixture models, maximum likelihood, spatial random-effects models.