Title of the tutorial

Statistical Learning for Multi-Sensor Signals and Multivariate Time-Series Data

Presenter

Dr. Pavel Loskot

Contact information

Address: ZJU-UIUC Institute, 718 East Haizhou Road, Haining, Zhejiang 314400, China
Phone: +86 571 8757 2579
Email: pavelloskot@intl.zju.edu.cn

Short biography

Pavel Loskot joined the ZJU-UIUC Institute in January 2021 as an Associate Professor after being 14 years at Swansea University in the UK. He received his PhD degree in Wireless Communications from the University of Alberta in Canada, and the MSc and BSc degrees in Radioelectronics and Biomedical Electronics, respectively, from the Czech Technical University of Prague in the Czech Republic. He is the Senior Member of the IEEE, Fellow of the Higher Education Academy in the UK, and the Recognized Research Supervisor of the UK Council for Graduate Education. His current research interests involve mathematical modeling, statistical signal processing and classical machine learning for multi-sensor data in biomedicine and in intelligent transport.

Selected publications of the presenter

- P. Loskot, “Discovering Causality in Event Time-Series,” *In Proc. 7th International Conference on Advances in Signal*, Image and Video Processing, Venice, Italy, 22-26 May 2022.
- T. Zheng and P. Loskot, “Signal Folding for Efficient Classification of Near-Cyclostationary Biological Signals,” *Mathematics*, vol. 10, no: 2:192, January 2022. DOI:10.3390/math10020192
- P. Loskot, “A Generative Model for Correlated Graph Signals,” *Mathematics*, vol. 9, no. 23:3078, November 2021. DOI: 10.3390/math9233078
- P. Loskot, “Polynomial Representations of High-Dimensional Observations of Random Processes,” *Mathematics*, vol. 9, no. 2:123, January 2021. DOI: 10.3390/math9020123
- P. Loskot, K. Atitey, L. Mihaylova, “Comprehensive Review of Models and Methods for Inferences in Bio-Chemical Reaction Networks,” *Frontiers in Genetics*, Vol. 10, No. 549, published 14 June 2019. DOI: 10.3389/fgene.2019.00549
- M. A. M. Hassanien and P. Loskot, "Non-Ergodic Error Rate Analysis of Finite Length Received Sequences," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3452-3457, September 2013. DOI: 10.1109/TVT.2013.2254141

Abstract

The major investments and subsequent progress in sensor technologies have led to mass deployment of different types of sensors in many different environments. Most of these sensors generate one-dimensional analog signals, which are then digitally processed. The challenge is that the real-world sensor signals are often non-Gaussian, non-Markovian and non-stationary. Moreover, the measurements can be sparse and across many dimensions, and the underlying data models are non-linear. In many

scenarios, multiple heterogeneous sensors are combined to create compounded views of the observed systems. The processing of sensor signals can be partially distributed to optimally utilize the available computing, storage and communication resources. Some applications may also require a real-time processing of the streaming data under the maximum latency constraints. This tutorial will discuss the recent progress in multi-sensor signal and multivariate time-series data processing. It is crucial to clearly formulate the data processing tasks and objectives with appropriate constraints, and to obtain the relevant data models before considering the current methods for processing multivariate time-series data including anomaly detection, spatial-temporal extrapolation, context learning, Bayesian machine learning, adaptive and semi-supervise learning, active learning with automated data labeling, distributed learning, learning at scale and other such tasks and methods.

Duration[*]

half-day (3 hours)

[*] The tutorial can be easily extended into a full-day (2 x 3 hours), if desired.

Importance and timeliness

The sensors are used very extensively to monitor various systems in engineering as well as in biomedicine. The use of multiple heterogeneous sensors is nowadays common in building smart and intelligent systems and environments. These new applications bring new requirements for online and off-line processing of the streaming multivariate time-series data to fully exploit their spatial-temporal dependencies and causal associations. Moreover, the generated sensor signals are often non-Gaussian, non-stationary and non-Markovian, and their models are non-linear and may be subject to non-additive distortions. This tutorial will cover first the key data acquisition problems including pre-processing of multiple interrelated analog sensor signals to aid the subsequent statistical learning (25%), and discuss the underlying data models and associated problems (25%) before concentrating on the main topic of the tutorial, i.e., the current methods of statistical learning for multivariate time-series data (50%). The tutorial will inspire new as well as existing graduate students and other researchers who are looking for challenging, real-world research problems in signal processing and machine learning, especially in the areas from processing physiological signals in biomedicine to designing real-time intelligent systems and the corresponding applications.

Outline and the material to be presented

The tutorial will focus on the recent developments in processing multivariate time-series data. It will be delivered in the following 3 parts with the most emphasis on the last, third part.

1. Data acquisition (45 min): sampling and decomposition of random signals, variance and dimensionality reduction methods, scaling and normalizations, spatial-temporal extrapolations
2. Data modeling (45 min): linear and non-linear high-dimensional models, models for non-stationary and/or non-Markovian signals, sparse and graph signals, signal topology and geometric interpretation, probabilistic and hierarchical models, structural causal models, model selection and dealing with modeling uncertainty
3. Statistical learning (90 min): statistical and causal inference including parameter estimation, anomaly and change detection, data fusion, data selection and labeling, unbalanced data

classes, adaptive semi-supervised learning, online learning, distributed learning, and learning at scale

Previous tutorial experience and the audience size of each past tutorial

The presenter has over 15 years of experience in higher education teaching, and had delivered over 10 tutorials in various international conferences. Specifically, the tutorials by the presenter in the past 5 years:

- "Network Metrics and Measurements in the Era of the Digital Economies," APSIPA ASC 2017, Kuala Lumpur, Malaysia, 12 December 2017. (about 15 attendees)
- "Internet Access in Challenging Environments" International Symposium on Communications and Information Technologies (ISCIT), Bangkok, Thailand, 26-29 September 2018. (about 25 attendees)
- "Geolocation Enhanced Mobile Communication Networks," Tutorial accredited for DoD Continuing Education by CompTIA CEUs as A+, Network+, and, Cloud+ certifications under credentials #8570.01-M, IEEE Military Communication Conference (MILCOM), Los Angeles, USA, 29-31 October 2018. (nearly 60 attendees)
- "Real-Time Aspets of Mobile Communication Networks" Tutorial accredited for DoD Continuing Education by CompTIA CEUs and GIAC CPEs as A+, Network+, and, Cloud+ certifications under credentials #8570.01-M, IEEE Military Communication Conference (MILCOM), Norfolk, VA, USA, 12-14 November 2019. (about 40 attendees)
- "In silico experiment design for hypothesis testing and generation," APSIPA ASC, Tokyo, Japan, 14 December 2021. (about 15 attendees)
- "Current Approaches to Graph Signal Processing," 6th International Conference on Advances in Signal, Image and Video Processing (SIGNAL), Valencia, Spain, 30 May – 3 June 2021. (about 20 attendees)
- "Processing Linear Graph Signals," 7th International Conference on Advances in Signal, Image and Video Processing (SIGNAL), Venice, Italy, 22-26 May 2022. (about 40 attendees)