

**Journée Statistique et
Informatique pour la Science
des Données à Paris Saclay**

Report of Contributions

Contribution ID: 1

Type: **not specified**

Information-Theoretic Methods in Data Sciences: Model Uncertainty, Robustness and Model Drift

Wednesday, January 26, 2022 10:20 AM (40 minutes)

Deep learning models are known to be bad at signalling failure: These probabilistic models tend to make predictions with high confidence, and this is problematic in real-world applications to critical systems such as healthcare, self-driving cars, among others, where there are considerable safety implications, or where there are discrepancies between the training data and data at testing time that the model makes predictions on. There is a pressing need both for understanding when models predictions should (or should not) be trusted, detecting out-of-distribution examples, and in improving model robustness to adversarial and natural changes in the data. In this talk, we will give an overview of those fundamental problems and key tasks. Namely, we first examine model uncertainty and calibration, and then we discuss simple but still effective methods for detecting misclassification errors and out-of-distribution examples, and for improving robustness in deep learning. We will describe information-theoretic concepts from fundamentals to state-of-the-art approaches, by going into a deep dive into promising avenues and will close by highlighting open challenges in the field.

Presenter: Prof. PIANTANIDA, Pablo (L2S/CentraleSupélec)

Contribution ID: 2

Type: **not specified**

Bi-level Optimisation for Machine Learning

Wednesday, January 26, 2022 11:10 AM (40 minutes)

In recent years, bi-level optimization – solving an optimization problem that depends on the results of another optimization problem – has raised much interest in the machine learning community. This type of problem arises in many different fields, ranging from hyper-parameter optimization and data-augmentation to dictionary learning. A core question for such a problem is the estimation of the gradient when the inner problem is not solved exactly. While some fundamental results exist, there is still a gap between what is used in practice and our understanding of the theoretical behavior of such problems. In this talk, I will review different use cases where this type of problem arises as well as recent advances on how to solve them efficiently.

Presenter: Prof. MOREAU, Thomas (INRIA Paris-Saclay)

Contribution ID: 3

Type: **not specified**

Machine Learning Competitions: a Meta-Learning Perspective

Wednesday, January 26, 2022 2:00 PM (40 minutes)

Our research aims at reducing the need for human expertise in the implementation of pattern recognition and modeling algorithms, including Deep Learning, in various fields of application (medicine, engineering, social sciences, physics), using multiple modalities (images, videos, text, time series, questionnaires). To that end, we organize scientific competitions (or challenges) in Automated Machine Learning (AutoML) and expose the community to progressively harder and more diverse settings, ever-reducing the need for human intervention in the modeling process. The code of winning teams is open-sourced.

In this presentation, we adopt the perspective that every challenge has a secret goal: that the winning algorithm will meta-generalize, i.e. perform well on new tasks it has never seen before. In particular, AutoML challenges, which test participants on multiple-tasks, can be thought of as meta-learning devices, aiming at training algorithms to perform well on tasks drawn from a particular domain, such that they will perform well in the future on similar tasks. Taking that angle, we apply the same principles of learning theory used to harness overfitting at the “regular learning level” to explain how to select a winner without meta-overfitting the tasks of the challenge. We will end with tips on how to organize your own challenge to further your own goals, and effectively meta-generalize!

Presenter: Prof. GUYON, Isabelle (LISN/INRIA Tau)

Contribution ID: 4

Type: **not specified**

Optimal Transport on Graph Data : Barycenters and Dictionary Learning

Wednesday, January 26, 2022 11:50 AM (40 minutes)

In recent years the Optimal Transport (OT) based Gromov-Wasserstein (GW) divergence has been investigated as a similarity measure between structured data expressed as distributions typically lying in different metric spaces, such as graphs with arbitrary sizes. In this talk, we will address the optimization problem inherent in the computation of GW and some of its recent extensions, namely the Entropic and the Fused GW divergences. Next we will illustrate how these OT problems can be used to model graph data in learning scenarios such as graph compression, clustering and classification. Finally we will present a novel approach performing linear dictionary learning on graphs datasets using GW as data fitting term which simultaneously provides convenient graphs modeling for the aforementioned applications and efficient approximations to the GW divergence.

Presenter: Prof. FLAMARY, Rémi (CMAP/Ecole polytechnique)

Contribution ID: 5

Type: **not specified**

Change-Point Detection in Dynamic Networks

Wednesday, January 26, 2022 2:50 PM (40 minutes)

Structural changes occur in dynamic networks quite frequently and its detection is an important question in many applications. In this talk we consider the problem of change point detection at a temporal sequence of partially observed networks. The goal is to test whether there is a change in the network parameters. Our approach is based on the Matrix CUSUM test statistic and allows growing size of networks. We propose a new test and show that it is minimax optimal and robust to missing links.

Presenter: Prof. KLOPP, Olga (CREST/ESSEC business school)

Contribution ID: 6

Type: **not specified**

Deep Learning Strategies for SAR Image Restoration

Wednesday, January 26, 2022 3:30 PM (40 minutes)

SAR (Synthetic Aperture Radar) images are invaluable data for earth observation. They can be acquired at any time, regardless of the meteorological conditions, and provide information on the characteristics of the earth, its height and its possible movement thanks to the phase information of the backscattered electro-magnetic field. Due to the coherent imaging of the SAR sensors, images present strong fluctuations due to the speckle phenomenon. This phenomenon is a major obstacle for the analysis and understanding of SAR images. After an introduction to SAR imaging and SAR data statistics, the objective of this talk is to present some deep learning strategies to restore SAR images, in particular plug-and-play techniques, supervised, semi-supervised, and self-supervised methods. We will show how introducing the model of speckle physics inside the deep learning framework allow to outperform the state of the art methods.

Presenter: Prof. TUPIN, Florence (LTCI/Télécom Paris)

Contribution ID: 7

Type: **not specified**

Closing words

Wednesday, January 26, 2022 4:10 PM (10 minutes)