Internship

Optimal Transport for Anomaly Detection and Localization

Internship Description

Optimal transport (OT) [1] is a powerful framework to define and compute distances between distributions (a.k.a. Wasserstein or earth mover's distance), with a tractable computation thanks to the Sinkhorn algorithm, of which an online version has recently been proposed [2]. Beyond, OT allows to exhibit the transport map between the distributions.
In this internship, we envision leveraging the OT theory to design algorithms dealing with out-of-distribution detection in a non-parametric setting that operates over sliding windows on time series. Specifically, we will target in an online manner the localization of abnormal samples.

Even at low rates, detecting and localizing efficiently abnormal situations can be paramount. The goal of this internship is to spot the abnormal samples from distributions. While computing the discrepancy between distributions with OT may only assess how close these distributions are, a detailed assignment information resides in the transport (coupling) map. The intern will study how the assignment resulting from partial OT, which transports only a given fraction α of the total probability mass [3], can be used in the out-of-distribution and outliers scenarios. Specifically, abnormal samples can be located from the map by relying on the position of the non-transported mass. Thereon, the intern will design statistical tests allowing to estimate the proportion of α of out-of-distribution samples. For that, she/he will investigate randomization for varying values of α. The randomization will be achieved by a bootstrap procedure on the samples of compared sliding windows.

The objectives of the intern are the following:
1- Familiarize with the investigated OT framework
2- Explore OT for anomaly detection on toy data
2- Devise deep-learning framework for real data from well-known benchmarks
3- Evaluate the developed methods on real data from an industrial partner

This internship may lead to a PhD thesis.

Research Environment

This intern will conduct her/his research within the Machine Learning group in the LITIS Lab, under the supervision of Dr. Maxime Berar, Prof. Gilles Gasso, Dr. Fannia Pacheco and Prof. Paul Honeine. This internship will be conducted within a research project gathering 9 permanent researchers of the LITIS Lab and the intern will also interact with several PhD students and interns also working on deep anomaly detection for time series.

Applicant Profile

- Student in final year of Master or Engineering School, in applied mathematics, data science, artificial intelligence, or related fields.
- Strong skills in advanced statistics and Machine Learning
- Good programming skills in Python

Location

LITIS Lab, Université de Rouen Normandie, Saint Etienne du Rouvray (Rouen, France).

Terms

5 to 6 months, starting in February or March 2024.

Application

Applicants are invited to send their CV and grade transcripts by email to:
maxime.berar@univ-rouen.fr, gilles.gasso@insa-rouen.fr, fannia.pacheco@univ-rouen.fr, paul.honeine@univ-rouen.fr.

References

[1] G. Peyré, M. Cuturi, et al., "*Computational optimal transport: With application to data science*," Foundations and Trends® in Machine Learning, 2019
[2] A. Mensch and G. Peyré, "*Online sinkhorn: Optimal transport distances from sample streams*," in NeurIPS 2020
[3] L. Chapel, M. Z. Alaya, and G. Gasso, "*Partial Optimal Transport with Applications on Positive-Unlabeled Learning*," in NeurIPS, 2020.