Scientific Area	Synchrotron Light Applications		
Project Title	Machine Learning for the Real-Time Analysis of X-Ray Spectroscopic Data		
Recruiting Institution	European Synchrotron Radiation Facility (ESRF)		
PhD awarding Institution	Grenoble Alpes University	**PhD Duration**	36 Months
Supervisor/Institution	Marius Retegan / ESRF		
Co-Supervisor/Institution	Pieter Glatzel / ESRF		
Secondment(s)	ML and HPC learning with ENGAGE partners		

Project Description

X-ray spectroscopy is an indispensable tool for element-specific analysis in most fields of natural sciences. It gives access to a wealth of information regarding the structural parameters and electronic properties of materials. Currently, mapping the experimental data to the various properties of materials usually relies on manual processing, i.e., every single spectrum is interactively analyzed by the scientist. However, with the ever-increasing acquisition rates allowed by the emergence of the extremely brilliant light sources, such workflows will become unpractical.

The current project aims to develop machine learning techniques to make fast and accurate predictions of properties (e.g., coordination number, oxidation state, concentration, etc.) from spectroscopic data. This approach has the potential to dramatically increase the amount of information that is extracted from the data. The development of machine learning models for spectroscopy is currently dependent on synthetic (theoretical) data. Consequently, there are open questions regarding the quality of the predictions made using these models when experimental data is provided as input.

The project will bring much-needed automation of the data analysis workflows for spectroscopy. This will enable non-expert users to make better use of their data and to make the techniques more accessible to an even larger scientific community and especially industry.

Project Objectives

- Build large datasets of theoretical calculations and experimental data. The datasets will be used to train machine learning models.
- Automate the generation of structural models with desired properties/features (e.g., coordination number, oxidation state, etc.) needed during the training using the existing Python software.
- Set up and test machine learning models that can accurately infer properties from calculated or experimental data.
- Design and implement an autonomous experimental acquisition pipeline that leverages the predictions made using the machine learning models.
- Apply the automated pipelines for the characterization of metal catalysts, e.g., Cu-zeolites, Pt nanoparticles

Required Candidate Qualifications

- MSc (or equivalent) in physics, chemistry, computational science, or a related field.
- Knowledge of computational chemistry/physics software.
- Knowledge of a modern programming language (e.g., Python) and machine learning techniques is an asset.
- Excellent communication skills (oral and written) in English.