Project: Machine learning tools for noble gas geochronology data reduction

CSM CSCI370 SU25
USGS Denver Argon Geochronology Laboratory

Organization Background

The U.S. Geological Survey provides science for a changing world, which reflects and responds to society's continuously evolving needs. As the science arm of the Department of the Interior and one of the world's leading scientific institutions, the USGS brings an array of earth, water, biological, and mapping data and expertise to bear in support of decision-making on environmental, resource, and public safety issues.

Project Description

Noble gas isotopic measurements provide critical information for understanding the timing of many Earth systems and geologic processes, from studying the source and evolution of fluids in groundwater systems, to understanding the structure of the mantle, to determining the ages and timescales of processes like intrusion, volcanism, and the formation of mineral deposits that are critical to national security and economic development. Data reduction of isotopic measurements is fundamental to all such geochronology studies to convert the raw signals measured by the detectors of noble gas mass spectrometers into meaningful values that can be interpreted to have geologic significance. The first step in this data reduction process is to regress isotope signal evolutions (which "evolve" over time during analysis) and extrapolate to the time the noble gases were in equilibrium just prior to analysis, known as "time-zero" or "t0."

For this project, students will work with USGS Denver Argon Geochronology Laboratory (DAGL) personnel to develop a machine learning (ML) model that can suggest the best-fit regressions for isotope signal evolutions, and identify potential issues with them to flag problematic analyses for human review. The students are free to train the model in any software environment (e.g., using Python and Pytorch or Tensorflow), though they will help develop a mechanism for the model to run while embedded in a Java desktop application that is being developed by the DAGL for noble gas data reduction (which may drive the choice of training library).

The ML model will likely need to emit regression suggestions, potentially including recommendations for outlier detection or data trimming (e.g., to deal with problematic data), as well as various warning or error messages that would help flag problematic analyses for an analyst to review. The model may need to accommodate analyst preferences for the pool of allowable regressions and derived values (e.g., determining the uncertainty of interpolated or extrapolated values) and outlier detection (e.g., choice of outlier identifier and detection thresholds). Part of the project may also involve implementing and testing additional regressions and outlier identifiers that can be used by the model during inference, over what is currently planned for the Java data reduction app. In addition, the project may involve developing D3js visualizations that would help an analyst evaluate how well the regressions fit the data, and/or the confidence of the ML model predictions. The completed model will ideally be partly or fully integrated into the Java data reduction application.

Desired Skillset

- ***Required:*** Proficiency with a programming language commonly used for training ML

models, such as Python
- ***Desired:*** Experience building and training AI models with common libraries, such as Pytorch or Tensorflow
- ***Desired:*** Experience with Java and relevant libraries (e.g., DeepLeaning4J, NP4J, Python4J) for executing ML models in the JVM
- ***Optional:*** Experience creating D3js visualizations

Student Benefit

- Opportunity to develop highly valued and versatile skills with:
    - Designing, implementing, and training AI models
    - Integrating the model into an application for eventual use by scientists in government, academia, and industry
    - Developing custom visualizations using the popular D3js library
- Opportunity to work with government professionals on a topic of importance to national security and economic advancement

Work Location

The preferred work location is primarily remote, with weekly in-person meetings at Mines and/or the Denver Federal Center. Remote meetings will be via Microsoft Teams.

Intellectual Property

Any Intellectual Property (IP) developed during the Project will be owned by the either USGS, Colorado School of Mines, and (or) its students as determined by U.S. law. Notwithstanding, for all IP and derivative data created during the Project, USGS will receive a Government Use license, allowing the Government to reproduce, publish, modify, or otherwise use the IP. Government use may include providing the IP and derivative data to the public via USGS websites or public software repositories.

Non-Disclosure Agreement

An NDA is not required.

Contact Information

Cameron Mercer (cmmercer@usgs.gov)
Leah Morgan (lemorgan@usgs.gov)