National Aeronautics and
Space Administration

Large Scale Applications of Machine Learning using Remote Sensing for Building Agriculture Solutions

March 05, 12, & 19, 2024

Session 1: 10:00-11:30 or 14:00-15:30 **EST (UTC-5)**

Sessions 2 & 3: 10:00-11:30 or 14:00-15:30 **EDT (UTC-4)**

Remote sensing data is becoming crucial to solve some of the most important environmental problems, especially pertaining to agricultural applications and food security. Effectively working with this large data source requires different tools and processing, such as cloud computing and infrastructure. Participants will become familiar with data format and quality considerations, tools, and techniques to process remote sensing imagery at large scale from publicly available satellite sources, using cloud tools such as AWS S3, Databricks, and Parquet. Additionally, participants will learn how to analyze and train machine learning models for classification using this large source of data to solve environmental problems with a focus on agriculture. Participants will have a basic understanding of tools such as Pyspark, TensorFlow, and Uber H3. We hope that participants in this course will walk away with the skills and tools to train algorithms using satellite imagery to solve environmental problems anywhere on the planet.

Part 1: Data Preparation of Imagery for Large-Scale Machine Learning (ML) Modeling

Trainers: Sean McCartney, John Just, Erik Sorensen

- Identify different types of remote sensing data used in agricultural applications, such as irregular time series, Sentinel-2 tiling grid, Uber H3 hex grid, CDL, S3.
- Apply correct procedure to run Python code in Databricks Community to download and process remote sensing data at large scale (> 5GB) with cloud tools.

Part 2: Data Loaders for Training ML Models on Irregularly-Spaced Time-Series of Imagery

Trainers: Sean McCartney, John Just, Erik Sorensen

- Follow the process to set up a Tensorflow data loader that works with Parquet files to create a training pipeline suitable for training a model on large-scale data
- Perform steps to manipulate the imagery data stored in tables, normalize the values, and bucketize irregularly spaced time-series data to prep for modeling
- Follow steps to parallelize/prefetch preprocessing for fast training
- Apply the correct procedure to split time-series data into train/val/test sets to avoid data leakage.

ARSET empowers the global community through remote sensing training.

appliedsciences.nasa.gov/arset

Part 3: Training & Testing ML Models for Irregularly-Spaced Time Series of Imagery

Trainers: Sean McCartney, John Just, Erik Sorensen

- Perform the process to set up and train a 1-D convolutional neural network (CNN) model that learns to detect crop-type from a satellite image.
- Follow steps to monitor model performance during training and how to choose appropriate hyperparameter adjustments.
- Plot the model to validate performance after training.

ARSET empowers the global community through remote sensing training.