Mapping Changes from Grasslands to Arable Lands with Advance Machine Learning Models Implemented in H2O Library

Provided R script has been developed as part of Ph.D. project that is aimed at mapping changes from grasslands to arable lands. The main research goal was to develop advanced classification methodology with machine learning technologies. Such technologies are implemented in H2O library (https://docs.h2o.ai/h2o/latest-stable/h2o-docs/index.html) especially in its AutoML function. AutoMl function implements advanced machine learning algorithms that are not so common in the field of remote sensing. AutoML function allows inexperienced users to properly tune input parameters of these algorithms without deeper knowledge. Provided R script implements functions that are avalable in H2O library and allows its users to classify remote sensing data.

Installation

Mandatory requirement for correct running of provided R script is proper JAVA installation. At first, please make sure that JAVA is properly installed on your computer- necessary path of JAVA installation directory is set as enviromnment variable for your operating system.

Then proper R version must be installed and R must be able to recognise proper JAVA installation. If mandatory requirements are met, then installation of H2O library can begin:

https://docs.h2o.ai/h2o/latest-stable/h2o-docs/downloading.html#install-in-r

Together with H2O library other packages, that are required for spatial data manipulation in R is required: raster, rgdal, caret, rChoiceDialogs. It is recommended to install these packages erlier than H2O library. It can be done manually or with a single command in R console:

```
install.packages(c("raster", "rgdal", "caret", "rChoiceDialogs", dependencies=T)
```

Please make sure that you have not installed these packages yet, otherwise it might lead to abnormal behaviour.

Input Data and their Structure

Sample data for demonstration can be found here:

https://data.mendeley.com/datasets/cvjjntm8wt/1

There are two directories located in provided zip file: *Training Data* and *Image Data*. Traning data directory contains NDWI Time Series dataset calculated from Sentinel 2 satellite. Image data directory contains reference point shapefile with two classes in numeric format. Class number 1 are points that refer to changes from grasslands to arable lands, class number 2 is connected wirth no change thematic class.

Outputs and Implemented Functions

- Land Cover Classification in the form of thematic raster
- ROC Curve Plot
- Producer's and User's and accuracies in the .txt file format
- Model Id Protocol provided by AutoML function in txt file format
- Confusion Matrix in .txt file format
- Traning and Validation Points in shapefile format
- ,,Trained Model" directory – directory contains MOJO files of all models tuned by AutMl function of H2O library: https://docs.h2o.ai/h2o/latest-stable/h2o-docs/mojo-quickstart.html