

Supervised Classification of Satellite Imagery Using Conventional Feature Extraction Methods

Image Interpretation

Assignment 1

Janik Baumer

Andreas Brown

Fergus Dal

27.10.2021

Introduction

The objective of this assignment was to classify satellite images using basic machine learning techniques implemented with well-known libraries in Python. Given a set of labelled Sentinel-2 images, a classifier was trained to predict certain labels or categories, namely palm oil plants, clouds, and the background (neither vegetation, nor cloud). There are countless techniques available to achieve this task including neural networks and deep learning methods. Our group focussed on more traditional and conventional methods of classification based on feature extraction. Once our model had been trained, validation and testing were carried out and finally, the quality of the model was assessed with various metrics derived from a confusion matrix. This report first describes our classification methods in more detail and the features that were extracted prior to presenting our results and interpretations.

Methods

The classification process can typically be subdivided into several sections or task groups, the first of which is data loading and pre-processing. Given the volume of image data, a special data loader function was needed to load batches of an image in order to avoid memory-overflow errors. During this stage, it was also important to establish a ground truth and remove pixels without data, visually noticeable as black areas not captured by the satellite. Note that atmospheric corrections are also essential and should be performed but these were already accounted for in the provided data. Several feature types were then defined, which would in turn be used as input feature layers to train a machine learning model. The features considered are briefly outlined below.

Features

* **NDVI:** The Normalized Differential Vegetation Index is very popular vegetation index used in remote sensing that is calculated for any image pixel according to the following ratio:

where is the intensity of the infrared channel and is the intensity of the red colour channel. The resulting values range from -1 to +1. Negative values indicate water and weather phenomena (clouds, rain, snow, etc.), whereas positive values denote vegetation. A small positive value close to zero is associate with dry, barren land, rocks and soil, or unhealthy vegetation. On the other hand, large positive values represent dense, healthy vegetation. Many other vegetation indices exist in literature but NDVI was specifically chosen given its widespread popularity, and the fact that we had the necessary channels to compute it.

* Window Level Transfer Function: Commonly used in biomedical image processing, the window level technique is essentially a mapping function that redistributes image intensity values to improve visibility of a particular feature.
* Effectively scales and transforms the intensities within a certain range to enhance features
* Gaussian Blur Filter

This is useful for reducing noise and remove artifacts from the image. Has a smoothing effect

* Mean of Nearest Neighbour
* Edge Detection (Canny)

Models

* K-nearest neighbour with varying k
* Decision trees
* Random forests
* Support Vector Machines (SVM)

Metrics

* Kappa Statistic
* Accuracy
* Precision
* Recall
* F1

Conclusion

Best method and recommendations

References

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540547/> (window-level)

<https://www.ospo.noaa.gov/Products/land/mgvi/NDVI.html> (NDVI values)

<https://www.usgs.gov/core-science-systems/eros/phenology/science/ndvi-foundation-remote-sensing-phenology?qt-science_center_objects=0#qt-science_center_objects>