Learning to Classify DSTL Satellite Images

Ting Wang, Master of Data Science Capstone Project
Department of Computer Science, University of Minnesota

Abstract

Satellite image is a valuable resource for us to better understand our planet. It has been applied to help us achieve better resource mobilization during disasters and efficiently monitor agricultural crops, forest, fire, global temperature at various time period and spatial location. However, one of the challenges of utilizing this advanced technology is object identification on the satellite image such as building, waterway, roadway etc that relies heavily on labeling object by hand or inefficient semi-automation. In this project, we proposed an end to end Extraction, Transformation and Loading pipeline to address satellite image processing, temperature at various time period and spatial location. However, one of the frequent used algorithm for segmentation is through aggregating the pixel level features and segmenting the features to polygons using graph-based region merging algorithm (Felz-Hut) [10] algorithm.

Learning Method

1. Boosting Decision Tree algorithm, an algorithm previously used to support robustness classifier training against overlapping class distributions and mislabeling of the training data was used at pixel level classification.
2. Object-based image analysis which employs segmentation and classification processes was explored at object-based image classification.
3. 2 cluster the similar pixels into segment polygon and aggregate the feature information with the pixels in this segment polygon using graph-based region merging algorithm (Felz-Hut) [10] algorithm.

Image Object Identification

- Masks are used to select pixels associated to specific object classes.
- Sixteen-band: contain wide spectral channels.
  - The 8 Multispectral bands from 400 nm to 1040 nm (red, red-edge, coastal, blue, green, yellow, near-RIR and near-RIG) and 8 SWIR spectral bands.
  - The 16 bands are distributed in the 3 different images with file ended with a .AAP in GeoTIFF format.

Problem

The image datasets consists of 16 band images (1km * 1km, approximately 3400/3400 pixels) which contain spectral information by capturing wider wavelength channels that include 8 Multispectral (red, red-edge, coastal, blue, green, yellow, near-RIR and near-RIG) and 8 SWIR spectral bands. This dataset is generated by merging polygons from images of 1.24m spatial resolution of 1.24m.

Learning to Classify DSTL Satellite Images