Classification of Soil Using Multi-Spectral Imagery From the Landsat Satellite

Alexander Leven*, Student, Purdue University School of Aeronautics and Astronautics*
Dante Sanaei*, Student, Purdue University School of Aeronautics and Astronautics*

Prof. Veeraraghava Raju Hasti, *Project Advisor, Purdue University School of Mechanical Engineering*

PROBLEM DEFINITION

In the design of observation satellites, one of the key fields of interest is in remote sensing. To summarize, remote sensing essentially allows for the acquisition of information on the surface of the Earth without making any in-situ physical measurements [1]. Using reflected solar radiation propagated from a location of interest, a satellite's detector is able to detect and analyze such back-scattered data. Such applications of remote sensing include, among numerous other examples, the ability for satellites to monitor meteorological changes, the atmospheric environment, or topographic conditions.

In this project, the team is using Landsat data obtained by the Australian Centre for Remote Sensing. The Landsat satellite is capable of obtaining a significant amount of information based on a specific imaged scene [2]. Here, the satellite data is taken from four digital images of the same area using different spectral bands. From this data, individual pixels have been labeled with a particular soil characteristic. The purpose of the project, then, is to analyze the dataset and classify a given pixel with the likely type of soil that is represented.

For this project, an artificial intelligence algorithm was selected to analyze the data because this method is the optimal way to effectively classify individual data points. From observing the dataset alone, one can see a string of 27 pixel values. As such, it would be impractical to use pure visual inspection to determine what type of soil sample these values corresponded to. Moreover, the large number of instances in the dataset does not show any clear correlation between these values and the necessary result (when just viewing the numbers). With an artificial intelligence, the inherent relationships between these pixel values and the appropriate classification would become apparent after running the proposed algorithms. Only through an AI would it be possible for users to classify such input data.

ASSUMPTIONS

It is assumed that none of the pixels in the dataset are taken on the boundary of the original image which may skew the accuracy of the results. Furthermore, it is also assumed that the actual classification of the training dataset has been accurately done (as it was done by a person on the physical site).

DESCRIPTION OF DATASET

The database consists of multi-spectral values of pixels obtained from a small area of an imaged scene. Such pixels are further derived from a 3x3 neighborhood within this satellite image. Also, recall that each pixel within the database corresponds to a value within 0 to 255. Moreover, for each row of pixels, there is a classification label associated with the central pixel of each 3x3 neighbourhood. Then for each neighborhood of 9 pixels, there are four spectral bands to yield a total of 37 columns in the dataset (36 pixels and 1 classification label).

To restate the problem description, the goal now is that given these 36 pixels to predict the classification number. To start, the given dataset has already been split into training and testing sets: training composed of 4435 instances and testing composed of 2000 instances. Within the dataset, there are six different types of soil classes -- the labels of which are shown below:

Table 1. Classification Labels

Label	Class
1	Red Soil
2	Cotton Crop
3	Grey Soil
4	Damp Grey Soil
5	Soil w/ Vegetation Stubble
7	Very Damp Grey Soil

PROPOSED METHODOLOGY

Using logical regression, support vector machines, Naive Bayes, and KNN, each instance will be classified in one of these four methods. As mentioned, the dataset consists of 36 numerical pixel values and a classification label. Thus, the AI model should be trained using these pixel values to be able to classify the type of soil. After the relevant analysis, the methods will be compared to determine the respective accuracy of each and find which method is most suitable.

Feasibility and Project Goals

The overall project is relatively feasible and can realistically be completed within the six week period. The following timeline displays the relevant dates when certain tasks should be completed.

Nov. 4: Finalize methodology for analyzing given dataset

Nov. 10: Produce Machine Learning model for given dataset

Nov. 14: Test finalized model and produce accurate outputs

Nov. 18: Analyze results and perform more background research

Nov. 30: Finish Report and Video

Relevance to Society

Remote sensing is a major component of orbital scientific research in that it allows for large surveys of areas to be analyzed. Some physical surface areas may be impractical to physically travel to or it may be simply infeasible for any physical analysis to be done. Thus, remote sensing allows such analysis to happen using satellite imagery -- the development of machine learning only helps to efficiently examine such data.

Using these image classification algorithms, pixels can be automatically grouped given new datasets [3]. When the results of such algorithms are combined with existing geographical information, there is the potential for significant results to be obtained from such relationships. For example, if topological data is combined with the soil type algorithm, then it is possible to find a relationship and potentially predict which areas on Earth would likely have a respective soil type. The relationships obtained from machine learning and remote sensing are invaluable and can provide immense benefit for understanding the Earth on a detailed scale.

References:

[1] Schowengerdt, Robert A. (2007). Remote sensing: models and methods for image processing (3rd ed.). Academic Press. p. 2.

[2] Statlog (Landsat Satellite) Data Set. University of California, Irvine Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/Statlog+%28Landsat+Satellite%29

[3] What is Remote Sensing? NASA EarthData. https://earthdata.nasa.gov/learn/remote-sensing#data-processing-interpretation-and-analysis