

Project Milestone

Zach Stecher

Due: 11/15/16

Abstract

I discuss the problem of solving image-based classification using readily available machine learning techniques. Specifically, I discuss the issue of accurately classifying 99 species of plants using binary leaf images with multiple features. For this project I compare the results achieved by different available algorithms to find the most effective solution, and discuss real world applications and ease of practical use. The three initial algorithms I would like to analyze are the Multi-Layered Perceptron(MLP), TensorFlow Neural Network, and Random Forest.

1 Introduction

The focus of this project is the classification of provided leaf images and corresponding data into one of a possible 99 species of plants (<https://www.kaggle.com/c/leaf-classification>). While this is generally regarded as a simple problem to solve using existing standard techniques, there is always opportunity to expand our understanding of existing problems and techniques by applying and comparing them. For this project I aim to compare a number of different classifiers for this problem to derive which can achieve the most accurate results, and the speed of calculation for each one.

1.1 Practical Application

Automated plant recognition as a subset of general image based classification has a multitude of real world applications. Beyond just the academic implications, an accurate classifier open to the public could aid in applications like medicinal research, species tracking, agricultural research, medicinal practice in low-income countries where manufactured pharmaceutical solutions are not available, and much more.

1.2 Milestone

So far I have downloaded and begun looking through the data as well as started applying it to some of the classifiers I mentioned earlier. I've tried it with a Random Forest implementation to get a sense for how it works and plan to make adjustments to try and achieve a better score. I've also attempted to download the TensorFlow environment, however it appears to not be available on Windows so I'll need to try it on Linux and see if the system can handle the load of data. I plan to also try the MLP code we used for class, and then my own implementation of it if time permits and see if there's a way to achieve a perfect score with it. Currently the competition leader has a score of 0 and my goal is to match that, or at least get as close as possible.

1.3 Moving Forward

My next steps are to finish downloading and implementing the first run of each of these three classifiers to get a baseline for all three. Following that there are two articles of particular interest I found from JMLR that I will finish reading through as I believe the ideas discussed within could be useful to this project. If it's possible to apply them, I will. Finally, I will tinker with the code on the classifiers to attempt to achieve a 90 percent accuracy or better with each one.