

Project Proposal

John Randis

9/22/16

For my project, I intend to do an application project involving the classification of leaves. This project involves building a classifier to identify and properly classify 99 species of plants. The benefits and implications of automated plant classification are far and wide reaching. For instance, applications include plant-based population tracking, crop and food supply management, as well as medicinal applications. Classification done by humans is often inaccurate, and leads to duplicate classifications. This is an area that machine learning can really benefit.

The UCI machine learning repository provides about 1500 plant images on Kaggle as sample data. They also provide pre-extracted features for you, for instance, the size, the shape, the texture and the color. As an early goal, Kaggle suggests to build a classifier based on the pre-extracted features provided for you, and then afterwards build your own set of features thereafter. I have set deadlines for myself to try and find the best way to go about solving this problem

Deadline 1: Research machine learning strategies and algorithms that may help with this problem.

Deadline 2: Apply these machine learning strategies in a way that will help to automate the classification of plants.

Deadline 3: Produce a working python script that accurately classifies the plants based on the pre-extracted features provided by the UCI repository.

Deadline 4: Create my own image scraper that will extract features from the binary leaf images.

Deadline 5: Run the python script with the new image features that have been extracted to see how it compares, and adjust mistakes accordingly.

With machine learning properly applied to this problem, I believe leaf classification can become fully automated and provide many benefits. One possibly library I have begun to look into is Keras, which is a neural network library for Python. I believe this will help the machine to start to recognize patterns and classify the images faster. As image scanning is a technique Ive wanted to experiment with but havent gotten the chance to, Im looking forward to beginning the project and applying machine learning to it.