Aerobotics

Data Science Project

Blueberry forecasting challenge problem

Background

Overview

Within the precision agriculture space, blueberry yield forecasting presents a unique and challenging problem, due to the complex growth dynamics. Typical crops, such as oranges, exhibit a single stage of growth, more or less, as they progress through the season. A single blueberry bush, on the other hand, may exhibit all stages of growth at any given point in time, once flowering has taken place. Forecasting blueberry growth is an immensely valuable capability, since it allows farmers to efficiently plan market and labour logistics. Conventional methods require enormous numbers of laborers to march through the fields and collect growth statistics by hand, every single week. At Aerobotics, we are uniquely positioned to tackle this problem in a fully end-to-end manner using our hardware and software suite. Specifically, through multiple flights over a blueberry farm throughout the season, we are able to build up a time-series of growth statistics and thereby forecast growth many weeks into the future.

Blueberry growth

Blueberries go through 6 primary stages, summarised in the table below. Once the berry has transitioned into the *blue* stage, it is ready to be picked. Berries transition sequentially through each stage.

Order	Stage	Average time in stage (days)
1	Green	21
2	Colour break 1	18

3	Colour break 2	11
4	Pink	5
5	Cherry	3
6	Blue	0

In addition to daily growth stage transitions, blueberries also exhibit a long-term trend in their average *blue* stage weights. Typically *blue* weight declines as the season progresses and the bushes deplete their resources. [Note about blue blueberries getting picked, so the blue berries on the bush being removed and new ones are being added.]

Data description

The goal of blueberry forecasting is to determine what the yield in *kilograms* will be at a given point in the future. To this end, we have collected:

1. 22 weeks of stage counts. This is in the **count_data.csv** file
2. 22 weeks of average blue stage weights in grams. This is in the **weight_data.csv** file

You'll notice that for some weeks, there is data missing. This is due to logistical failure, and is bound to be encountered in a production setting. The table below gives a snapshot of 4 weeks worth of count data.

week	date	green	colour_break_1	colour_break_2	pink	cherry	blue
39	2018-09-27	660	327	65	19	29	54
40	2018-10-01	600	292	65	18	23	109
41	2018-10-08	532	240	85	25	18	65
42	2018-10-15	447	212	76	30	21	29

For example, the line for week 39 shows that there were 54 berries in the blue stage. At the end of this week, all 54 of those berries are picked, and their average weight is given in week 39 in the weight table. Similarly, the line for week 40 shows that 109 additional berries entered the blue stage from other stages and were subsequently picked.

Aerobotics

Task

The task is to come up with an approach to forecasting blue stage counts and weights 1, 2, 3 and 4 weeks into the future. That is, given N weeks of history, forecast the next K weeks of *blue* counts and a average weights. Note that logistically, the product is more robust if it can work with a minimal number of historical data points, and is more valuable the further ahead it can forecast. The performance of the solution must be quantified - the particular metric used is up to you, but requires justification.

Necessary Data Files

Link here

Technical requirements

- Solution must be coded in Python
- Solution must be packaged in a form that is easily runnable. We recommend building the solution into a docker container, or at the very least providing comprehensive documentation on package dependency.
- Solution must be clearly explained in a separate document

Communication

Please contact Michael at michaelm@aerobotics.com with any queries.