Cohort 12: 5th Meeting Minutes

21st of October 2022

Group: 90, Chair: Muchaveleli Manjate, Minutes: Devlan Mckenzie

P101 Benjamin Palay and Raphi Druion (Facial recognition)

Determined that manual machine learning wont work. Found a model that can be trained however the model struggles when zoomed in on features like the eye due to pixelation. To solve this problem, we took 9 points and averaged their values. Attempting to use the cluster to get 10 000 images for training. Proceeded to demonstrate a prototype which is trained using 700 images. A real time database has been implemented which tracks the number of iterations it took to get a similar face and creates a graph showing how closely related the features are. Trying to get a larger set to increase the accuracy of the program.

G74 Haffejee Mohammed and Thabo Tshabalala (App to scan pool water test strips)

Demonstrated their program which can scan test strips but is inaccurate. Over the next 2 weeks they plan to further increase the scan accuracy which is currently correct 1 out of 10 or maybe 20 times. Proceeds to further demonstrate the program showing features such as photo capture, new pool and auto processing.

Prof Estelle: Remember to document what you did and didn't manage to achieve for future work to flow smoothly. Additionally, the code must be accessible for the future work.

They found 2 methods to increase the accuracy of the program. The first method is to fix under and overexposed images. The next method is to generate colour charts and uses those to group accuracy. The plan for the coming week is to work on increasing the accuracy.

G75 Ahmed Ibrahim and Kevin Naidoo (The lost cities of South Africa)

Didn't bring a laptop and will demonstrate the program next week. They refined the SSD model which works acceptably now. Want to swap to an SNN model with increased accuracy and prof wants to get specific location images scanned. SNN will take longer to work with and analyse but will increase the overall accuracy of the program and they plan to work on it in the coming week.

G51 Robin Jonker and Tristan Lilford (Location aware scientific workflows)

Greetings, on Wednesday they were given workflows to work with and have been integrating the static and dynamic aspects of their project with the new workflows. Run time is 6m 30s approximately and 2m 18s for the optimal. The dynamic times are 4m 16s. Lately been looking into using AWS and spoke with the prof but its not looking like its going to happen. They have been doing lots of research into S3 backet. Currently they can pick a region and moved onto dummy tests and are in the process of moving into actual testing. In response to a question about reserving nodes to increase run time they said that they cannot reserve nodes as that would defeat the purpose of the project which is to dynamical assign nodes and determine which nodes would be best to use.

G48 Nkosingiphile Ndabandaba and Ishmael Sithole (Fast genotype calling)

Finalised the data which was generated from an array. Testing if the program can read the data and did 2 sets of testing which involved assuming different file sizes. They demonstrated the data set which the prof gave them. The first task was to determine the pathing of each file and to open each file. There are 4 genotypes using red and green colours. They spoke with prof and were

recommended to use integers to store the data and to place them in an csv file. Then use the csv file to plot the data. An issue they ran into is the lack of publicly available resources related to this project. So, they attempted to research the problem in greater depth to find resources. Looking into clustering the data as they require it. They have 2 aims in mind the first is to use parallelism to improve the code and the second is to use the prof's function to plot the red and green data as the 4 genotypes. An issue is that they are unable to cluster all the data together and are looking into a way to solve that.

G67 Johann Gouws and James Allsop - Mapping the Rainbow (twitter):

The main issue is getting followers as even a small set of data takes 3 hours to process. The prof suggested that they contact the API people to get their professional help but only managed to get a meeting for next Friday. They did however manage to improve the code so that it takes 1h 30 minutes but that is still too time consuming, and automation limits the control they have over the data. Looking into research material, which is centred around topics, search for topics and then search for tweets which is faster but gives a poor connection in terms of corelation. This change would alter the entire project scope, direction and dynamic. As the scope has basically changed to a topic basis, they plan to create a bounding box next week and pull topics to see sample size. If all else fails, they are just going to use topics to pull data because if they get sample data then they can apply a model to it and obtain a variety of information. They are aiming to get a MVP as there is simply too much information to filter through.

G36 Joseph Baggott and Hraklis Papageorgiou- The world's protest capital:

Last weeks issue with the API was resolved by pulling all the SA protests and creating their own database which is much faster to pull data from than before, however this process did take awhile due to the tweet analysis, spark, tactics, and location data collected in a broad sense. The program does sentimental analysis on a scale of 1 to 100.

Asked by prof Estelle to do sentimental analysis on the lightning storm which occurred during the week. And to dump the tweets for later use.

Look into vid analysis but determined that it's a bit beyond the scope of the current project.

G15 Rael Ware and Gia Croock- Cheap Air Quality Reporting Station:

Demonstrated the application which shows live data from the SDI location. The air is explained to be clean and for testing purposes they will go burn toast at the location after this meeting. The data is analysed in Matlab on an hourly and daily level. They proceed to discuss various air quality monitor costs and explain that their monitor can sense at a 2.5 level and a 1 level which is less accurate than the 2.5 level. The 1 level tends to overestimate and is thus less accurate, but the data is in the right shape and thus they plan to calibrate the measurement device to be more accurate. At the 2.5 level the device is very accurate and not much work is needed for this level. They explain that at 1 level it is 77% accurate and after processing they get 92% accuracy but discuss how external factors can affect this accuracy.

P61 Devlan Mckenzie and Muchaveleli Manjate- Network firewall management and visualisation:

They explained that they met with the prof and got further advice with regards to converting the rules into Boolean expression and creating the BDD. They have been working in tulip and control and have basic rule conversion working. There is still and issue with IP address conversion, but all other field can convert into a Boolean expression successfully. The group also started looking into using

Pyeda which is another python package which deals with Boolean expressions and BDD creation due to the technical difficulty of using the lower-level tulip and control packages. The basic level of BDD can be created and is seen to work but the group has not yet combined all the rule expression into a single BDD. The plan for the coming week is to further refine the logic around rule conversion and to store and create the entire ruleset as a BDD.

Meeting adjourned.