Carolyn Heinzer
05/31/24

Task 1: Validation Set

- Description:
  - I used the OS module to more easily check for, create, and remove files. First, I checked for Validation and Train folders and created them if they were not already present. Then, while a 0-indexed counter was less than 24, I used random.choice to randomly choose a file.
  - To make sure there would be no duplicates in the set, I only continued if the file was not in a list I added each file to after using it. If the file was not in the list, I renamed it, effectively moving it to the Validation folder, then added the file name to the used list and incremented the counter (if the file name is a duplicate, I did not).
  - I checked every file in Labeled and if not in used, I added it to the training folder using the same process as outlined above.
    - I would have created duplicate files, but this messed with the format of the csv, and I didn't want to spend inordinate amounts of time trying to figure out the problem. I instead just reverted the folders every time.
- Resources I used: https://www.w3schools.com/python/python_file_write.asp, https://www.w3schools.com/python/ref_file_writelines.asp, https://favtutor.com/blogs/move-file-python#:~:text=To%20relocate%20a%20file%20from,file%20in%20the%20file%20system.&text=It's%20vital%20to%20understand%20that%20the%20os, https://www.geeksforgeeks.org/os-module-python-examples/

Task 3: Shallow Learning Classifier

- Description:
  - First, in the predict_shallow_folder function, I extracted all useful attributes from each file, calculated some attributes of those points, and combined those into a summary file ("data_summary.csv"). I only chose three attributes – the x, y, and z components of velocity. These were the most similar to the accelerometer readings used in the CenceMe paper. I used the x, y, and z velocities from the left controller since the controller's movements are largely different for each activity as opposed to the headset's, which stays in place for standing and sitting, and barely moves for arm circles, arms stretching, and driving. I only considered the left controller because the right controller varies in a very similar way.
  - Also taking inspiration from the CenceMe paper, I calculated the mean, standard deviation, and number of peaks for each component of velocity and used those as features in a decision tree.

Carolyn Heinzer
05/31/24

    - I then created the decision tree by using Pandas to create a dataframe by reading from the combination file I created. I made sure that all data was numeric. I separated the feature and target columns (feature columns being x_peaks, x_mean, x_std, y_peaks, y_mean, y_std, z_peaks, z_mean, z_std, and the target column being code), and created an instance of a DecisionTreeClassifier and called fit on it using the feature and target columns. The result of the prediction was essentially an integer, so I converted it back into a three character code.
    - The pre-written code mapped the predict_shallow function, although I made a small edit to make sure the data_summary file was not included in that.
- Latency: about 30s
- My classification report may be found in classification_report.txt
    - Small note here – my results for accuracy and precision are suspiciously high. I may have made a mistake with their calculations or with my predictions, but I can't seem to find anything.
- Resources I used:
    - https://www.geeksforgeeks.org/os-module-python-examples/
    - https://www.geeksforgeeks.org/decision-tree-implementation-python/
    - https://medium.com/@stella96joshua/how-to-combine-multiple-csv-files-using-python-for-your-analysis-a88017c6ff9e
    - https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html
    - https://matplotlib.org/stable/tutorials/pyplot.html
    - https://www.geeksforgeeks.org/create-the-mean-and-standard-deviation-of-the-data-of-a-pandas-series/
    - https://www.geeksforgeeks.org/writing-csv-files-in-python/
    - https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html