Model Setup Guide

Python 3.8.10 is used with the python3 command on Linux for executing python files.

ML_CODE:

The model uses two files to run: model.py and setup.py.

The files assume a Data folder structure in the same directory as them following: [Data/Respeck/s1/s1_asce...] which contains the csv files for each student titled in the following way: [StudentNumber]_[Sensor]_[ActivityName]_[RespiratorySymptom].csv.

When setup.py is run it will pass through the Data folder and convert each students data into a tfrecord containing three second segments of each of the data files with the associated labels and pre-processing. We have included the pre-calculated tfrecords for simplicity but if you need to verify them then carry out the above steps.

Once all of these tfrecords have been created, the model.py file can be run from the same directory which will use the tfrecords to train and test the model. There is a function at the bottom called loo(n) which takes in the number of epochs per run and will perform loo testing and output a final score

Notes:

- Requirements.txt will say there is a missing tensorflow package, this is not a problem
- The running of the system assumed the files are in the same order as the training data was as the order of labels is important for ensuring the model works, they should be:
  Respiratory - ['other', 'breathingNormal', 'coughing', 'hyperventilating']
  Activity - ['ascending', 'descending', 'lyingBack', 'lyingLeft', 'lyingRight', 'lyingStomach', 'miscMovement', 'normalWalking', 'running', 'shuffleWalking', 'stationary']
- The last line is commented out but if uncommented will save the model as an h5 model
- The model does not reset between runs for LOO as a way of decreasing the number of epochs for each run, otherwise the model takes unnecessarily long to run. This has the same effect as heuristically choosing starting weights which is common practice in machine learning.

Model_evaluation

I have included necessary comments at the top of the file, most noticeably however is that only one model is used for all tasks and it does not filter data depending on which task it is doing in the script, and as such to test for a specific task it must exclusively be fed appropriate data for said task.