SPEECH COMMAND MODEL

February 5, 2021

- **OVERVIEW:**

  Speech recognition is the process of converting human sound signals into words or instructions. It is based on speech. It is an important research direction of speech signal processing and a branch of pattern recognition.Speech recognition applications include voice user interfaces such as voice dialing (e.g. "call home"), call routing (e.g. "I would like to make a collect call"), domestic appliance control, search key words (e.g. find a podcast where particular words were spoken), simple data entry (e.g., entering a credit card number). In similar way Speech voice recognition model is based on concepts of Convolution, LSTM , Attention and recognise pretrained voice with accuracy of **99.6%**.

- **DATA:**

  1: Set 16KHz as sampling rate.

  2: Record 80 utterances of each command.

  3: Save samples of each command in different folders.

* Speech/forward.

* Speech/back.

* Speech/left.

* Speech/right.

* Speech/stop.

- **Model Summary**

```
Layer (type)                 Output Shape          Param #     Connected to
===================================================================================
Input (InputLayer)           [(None, 49, 39, 1)]   0

Conv1 (Conv2D)               (None, 49, 39, 10)    60          Input[0][0]

BN1 (BatchNormalization)     (None, 49, 39, 10)    40          Conv1[0][0]

Conv2 (Conv2D)               (None, 49, 39, 1)     51          BN1[0][0]

BN2 (BatchNormalization)     (None, 49, 39, 1)     4           Conv2[0][0]

Squeeze (Reshape)            (None, 49, 39)        0           BN2[0][0]

LSTM_Sequences (LSTM)        (None, 49, 64)        26624       Squeeze[0][0]

FinalSequence (Lambda)       (None, 64)            0           LSTM_Sequences[0][0]

UnitImportance (Dense)       (None, 64)            4160        FinalSequence[0][0]

AttentionScores (Dot)        (None, 49)            0           UnitImportance[0][0]
                                                               LSTM_Sequences[0][0]

AttentionSoftmax (Softmax)   (None, 49)            0           AttentionScores[0][0]

AttentionVector (Dot)        (None, 64)            0           AttentionSoftmax[0][0]
                                                               LSTM_Sequences[0][0]

FC (Dense)                   (None, 32)            2080        AttentionVector[0][0]

Output (Dense)               (None, 5)             165         FC[0][0]
===================================================================================
```

- **The model is successfully built and has achieved the highest accuracy of 99.6%**

- **RUN:**

  The Code is written using Google Colab:

  1. Open ColabNotebook.ipynb and change Runtime to GPU.
  2. Upload Speech-Recognition/Speech to Colab.
  3. Change data-dir in all cells to point to Speech-Recognition/speech.
  4. Run the cells in the same order in Notebook Test.

- **TEST:**

  1: Locate the folder where you save your model.h5 file.

  2: Start speaking when you see mike in the bottom right pane of the task bar or see red blinking dot in the title bar.

- **Language Used:**

  PYTHON

- **Libraries and Packages Used:**

  KAPRE, SCIKIT LEARN, SOUND FILE, TENSORFLOW.