Desh Iyer | 500081889 | Experiment - 9

April 2, 2023

1 Experiment 9 - Speech Recognition

This is an experiment that highlights the use of `huggingface` transformers in Python making it extremely easy to perform speech recognition on audio files with just a few lines of code. This notebook will compare various pre-trained speech recognition models that are accessible through the aforementioned `transformers` library. This experiment uses sound files in `.wav` format from the TESS dataset.

1.1 Import Dependencies

Here we define the `transformers` library to import Facebook's `wav2vec2-base-960h` speech recognition transformer. Also, `pipeline` from `transformers` helps us pass all of our input `.wav` files to the model.

[ ]:
```
from transformers import pipeline
import pandas as pd
import warnings
import os

warnings.filterwarnings('ignore')
```

```
2023-04-02 23:42:32.382714: I tensorflow/core/platform/cpu_feature_guard.cc:193]
This TensorFlow binary is optimized with oneAPI Deep Neural Network Library
(oneDNN) to use the following CPU instructions in performance-critical
operations:  AVX2 FMA
To enable them in other operations, rebuild TensorFlow with the appropriate
compiler flags.
2023-04-02 23:42:32.787138: W
tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could
not load dynamic library 'libcudart.so.11.0'; dlerror: libcudart.so.11.0: cannot
open shared object file: No such file or directory
2023-04-02 23:42:32.787180: I
tensorflow/compiler/xla/stream_executor/cuda/cudart_stub.cc:29] Ignore above
cudart dlerror if you do not have a GPU set up on your machine.
2023-04-02 23:42:34.345686: W
tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could
not load dynamic library 'libnvinfer.so.7'; dlerror: libnvinfer.so.7: cannot
open shared object file: No such file or directory
```

```
2023-04-02 23:42:34.345925: W
tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could
not load dynamic library 'libnvinfer_plugin.so.7'; dlerror:
libnvinfer_plugin.so.7: cannot open shared object file: No such file or
directory
2023-04-02 23:42:34.345946: W
tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Cannot
dlopen some TensorRT libraries. If you would like to use Nvidia GPU with
TensorRT, please make sure the missing libraries mentioned above are installed
properly.
```

1.2 Transformers Example for Speech Recognition

1.2.1 Create Pipeline

Creating the speech recogniser's pipeline using `huggingface`. We pass whatever model we require as an argument to the `pipeline()` function. As you can clearly see, we're using the Facebook `wav2vec2-base-960h` model.

[ ]:
```
speechRecognizer = pipeline(task='automatic-speech-recognition',␣
↪model='facebook/wav2vec2-base-960h')
```

```
Some weights of Wav2Vec2ForCTC were not initialized from the model checkpoint at
facebook/wav2vec2-base-960h and are newly initialized:
['wav2vec2.masked_spec_embed']
You should probably TRAIN this model on a down-stream task to be able to use it
for predictions and inference.
```

We're only going to handle the 200 odd `.wav` files in the `OAF_Fear` directory. To do this, we make a list to store the paths of all the files in the directory.

[ ]:
```
fearFiles = []

for dirname, _, filenames in os.walk('/kaggle/input/
↪toronto-emotional-speech-set-tess/TESS Toronto emotional speech set data/
↪OAF_Fear'):
    for filename in filenames:
        fearFiles.append(os.path.join(dirname, filename))

len(fearFiles)
```

[ ]:
```
0
```

Once we have the file paths to all 200 to all of those files, we can start to apply our pipeline to each of those files. We then create a list of the output from each of those files being passed to the model pipeline.

1.2.2 Pass Data to Pipeline

[ ]:
```
predictedDicts = []

for file in fearFiles:
    predictedDicts.append(speechRecognizer(file))
```

All the outputs we have in the list are of the dictionary datatype. We now need to extract the text from the dictionary and we do that in the following way.

[ ]:
```
text = []

for dictionary in predictedDicts:
    text.append(dictionary['text'])

len(text)
```

[ ]:
```
0
```

Converting this list of text to a pandas dataframe to easily view the text extracted.

1.2.3 Representing Data in DataFrame

[ ]:
```
data = pd.DataFrame(data=text, columns=['Speech Text'])
data
```

[ ]:
```
Empty DataFrame
Columns: [Speech Text]
Index: []
```

[ ]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 0 entries
Data columns (total 1 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Speech Text  0 non-null      object
dtypes: object(1)
memory usage: 0.0+ bytes
```

[ ]:
```
data.describe()
```

[ ]:
```
       Speech Text
count            0
unique           0
top            NaN
freq           NaN
```