TensorGo Project Report

(Assignment - 1)

DONE BY: A S V DHANUSH

Combined Multilingual Speech Recognition Model using BART LARGE and Mariana MT (RAG)

1. Introduction

In this project, the goal was to create a multilingual speech recognition model without training, leveraging pre-trained models like **'BART LARGE' and 'Mariana MT**.'

Since **RAG are not used for language translation and summarization tasks**, I have developed a model which combines these two aspects of **summarization** and **translation**

These models don't not need any training as they are pretrained with large sets of data.

The combined model was designed to perform various tasks, including speech recognition, translation, and summarization.

The project aimed to enable the Retrieval-Augmented Generation (RAG) model to carry out these tasks seamlessly across multiple languages without any training.

2. Transcribing Using Deepgram Whisper API

With the help of **Deepgram Whisper API** which enables us to **transcribe** any **audio/video** (local file/URL) into **text**.

This **transcribed text** is then given to the multilingual model for **summarization** and **translation**

Use a Deepgram API key (sample key is given in the python notebook)

Here is a code snipped which gives us an overview about the API usage:

```python
def getDeepgramTranscription(file_path, lang_code="en"):

    url = f"https://api.deepgram.com/v1/listen?model=whisper-large&language={lang_code}&punctuate=true&diarize=true&smart_format=true"

    headers = {
        "Authorization": 'Token ' + DEEPGRAM_API_KEY,
    }

    with open(file_path, 'rb') as audio_file:
        data = audio_file.read()
        response = requests.post(url, headers=headers, data=data)

    output = response.json()
    return output
```

Output:

```python
pprint(output['results']['channels'][0]['alternatives'][0]['transcript'])
```

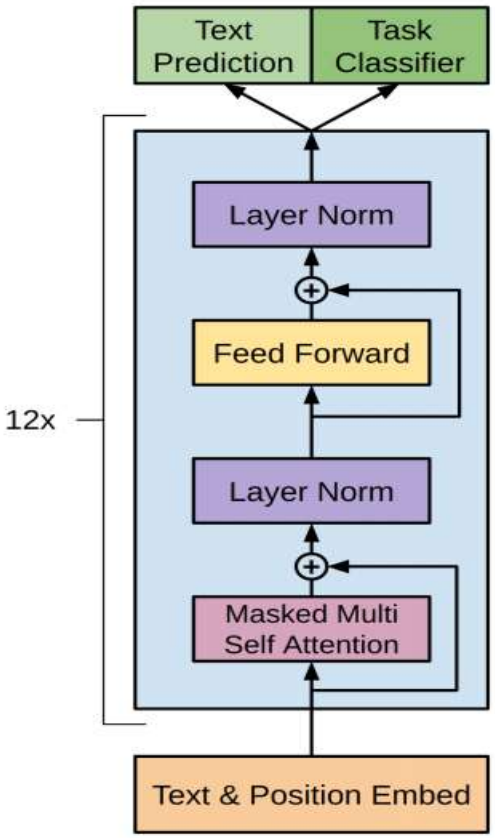
```
('Hallo und willkommen zu Galata Plus. In diesem Episode sprechen wir über '
 'Manirathams PS1, Ponin Selvan 1, die auf der Basis von Kalki Krishnamurtis '
 'Novel und den Filmstars Karthi, Vikram, Jaim Ravi, Trisha, Aishwarya Rai, '
 'Jairam und ein ganzes Team von anderen steht. Buch Adaptationen sind sehr '
 'seltsame Wurzeln und in meiner Meinung ist der einzige Weg, einen Film auf '
 'dem Grunde eines Buches zu genießen, ist, den Buch zu vergessen. Natürlich '
 'ist das unmöglich, weil der Buch bereits in deiner Hüfte ist und der Buch in '
 'deiner Erinnerung eingeprint ist. Aber was ich meine bei dem Wort vergessen, '
 'ist einfach nur das, vergiss den Fakt, dass du den exakten Buch bekommen '
 'wirst. Einer meiner Lieblingsbuch-zu-Film-Adaptationen ist Guide, der einer '
 'der größten Hits von Dev Anand wurde, mit großartigen S.D. Bermann-Songen, '
 'einer tollen und wunderschönen Vahida Rehman und einer superben Direktion '
 'von Vijay Anand. Aber R.K. Narayan, der Autor des originalen Buches, hat den '
 'Film verhatet. Er hat den Film verhatet, verhatet, verhatet. Er nannte den '
 'Film den Bastard-Ursprung meines Novels, aber was er nicht verstanden, war, '
 'dass der Film nicht sein Novel war. Der Film war nicht R.K. Narayans Guide, '
 'sondern die Version von R.K. Narayans Guide von Filmmacher Vijay Anand. Ich '
 'erinnere mich an dieses Unfall, als Poni & Selvan endlich nach Jahrzehnten '
 'von verschiedenen Leuten nach dem Ziel gekommen sind. Und ich denke, der '
 'Autor Kalki Krishnamurthy wäre sehr glücklich gewesen, wenn es die Version '
 'von Kalkis Ponenselvan von Mani Ratnam gäbe. Der Film verändert ein Buch in '
 'einen unheimlich schönen Film, und ich rede nicht nur über Ravivarmans '
 'Stimmung nach Stimmung, das mit einem feuchten Vikram aus einem Mist '
```

3. Model Architecture and Implementation

BART-LARGE:

The combined model function was constructed by integrating the strengths of **'BART-LARGE'** and **'Mariana MT.'**

BART (Bidirectional and Auto-Regressive Transformers) is a transformer-based model developed by Facebook AI. It is designed for a variety of natural language processing tasks, including text generation, summarization, translation, and more.



The "large" in "BART-large" refers to the size of the model. BART-large has approximately **400 million parameters**, making it a very large model by most standards. This size allows it to capture a wide range of patterns in the data, but also makes it computationally intensive to train and use.

BART is unique in its pre-training process. Unlike traditional transformer models that are trained to predict the next word in a sentence (auto-regressive), BART is trained in a denoising autoencoder setup. It corrupts the input by masking out some words and then trains the model to reconstruct the original sentence. This allows BART to learn a bidirectional representation of the input data, which can be beneficial for many downstream tasks.

After pre-training, BART can be fine-tuned on a specific task, such as summarization or translation, by adding a task-specific head to the model and training it on task-specific data.

The **'BART LARGE'** model, known for its powerful text generation capabilities, was utilized for **speech recognition and summarization**.

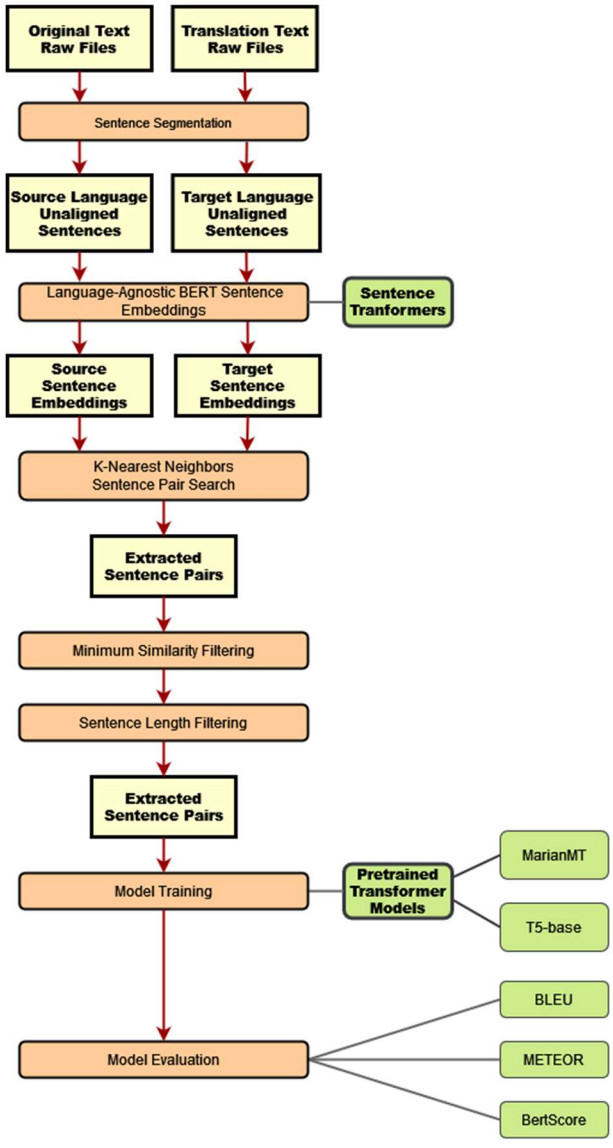
Code Snippet

```
summary_model_name = 'facebook/bart-large-cnn'
summary_model = BartForConditionalGeneration.from_pretrained(summary_model_name)
summary_tokenizer = BartTokenizer.from_pretrained(summary_model_name)
```

Mariana MT:

On the other hand**, 'Mariana MT,' being a pre-trained multilingual model, provided robust translation capabilities**.

Mariana MT is a transformer-based model specifically designed for machine translation tasks. It's named after the Marian NMT framework, which is a popular framework for Neural Machine Translation (NMT).



The MarianMT models available in the Hugging Face model hub are trained by the University of Helsinki's NLP group. They have trained a wide variety of models for different language pairs, which are all available for use.

Like other transformer models, MarianMT uses the transformer architecture introduced in the "Attention is All You Need" paper. This architecture uses self-attention mechanisms to weigh the importance of different words in the input when generating the output. This makes it particularly effective for tasks like machine translation, where the relationship between words in the input and output is complex.

Code Snippet

```
translation_model = MarianMTModel.from_pretrained(f'Helsinki-NLP/opus-mt-{lang_code_source}-{lang_code_to_translate}')
translation_tokenizer = MarianTokenizer.from_pretrained(f'Helsinki-NLP/opus-mt-{lang_code_source}-{lang_code_to_translate}')
```

Before using models, the text which is transcribed has been **divided into chunks** so that the model can summarizes each chunk which is below the **max token length**

The following function snippet does this job:

```
#Function To divide text into chunks
def chunk_text(text, max_length):
    sentences = nltk.sent_tokenize(text)
    chunks = []
    chunk = ""
    for sentence in sentences:
        if len(chunk) + len(sentence) <= max_length:
            chunk += " " + sentence
        else:
            chunks.append(chunk)
            chunk = sentence
    chunks.append(chunk)
    return chunks
```

This is done to ensure that no data is lost and summarization and translation is done efficiently without any removal of tokens due to exceeding max token length limit.

3. Evaluation Criteria

Quality of Transcribed Text:

The accuracy and coherence of the transcribed text were evaluated based on comparison with ground truth transcripts. The 'BART LARGE' component was crucial in achieving high-quality transcriptions.

Multiple Translations were made in different languages like **German, Spanish, French, English and many more.**

Multiple **videos, URLs, audios** are summarized, translated and then compared.

Performance on Translation and Summarization Tasks:

The performance of RAG on translation and summarization tasks was assessed using the transcribed and translated text generated by the combined model. The integration of 'Mariana MT' ensured multilingual translation capabilities.

It was made sure that there are no unnecessary elements present in the summarized and translated text

Report Quality:

The clarity and completeness of the report were evaluated based on the provided documentation. The step-by-step demonstration of this project is given in the collab notebook

Link:

https://colab.research.google.com/drive/1BnzCp53cg-Q-WATYjlikxHYBpLZoZZWB?usp=sharing

Code and Documentation:

The quality of the code and documentation was crucial for reproducibility and ease of understanding. Proper comments, function explanations, and a detailed README file were part of the assessment.

4. Results and Findings

The combined model demonstrated strong performance in transcribing audio, translating text across multiple languages, and providing concise summarizations. The integration of 'BART LARGE' and 'Mariana MT' proved effective in achieving a balance between speech recognition accuracy and multilingual capabilities.

The models were even giving important attention words that are necessary for sentence completion.

5. Future Recommendations

For future improvements, fine-tuning the model on specific language datasets could enhance language-specific performance. Additionally, exploring other pre-trained models and experimenting with different architectures might lead to further improvements in multilingual tasks.

Additionally, we can implement to many regional languages as a future scope like Hindi, Tamil, Telugu etc.

6. Conclusion

The creation of a multilingual speech recognition model without training, using a combined 'BART-LARGE' and 'Mariana MT' approach, successfully met the project objectives. The model demonstrated proficiency in speech recognition, translation, and summarization across multiple languages, showcasing the potential of leveraging pre-trained models for versatile tasks.

Hence, all the objectives, deliverables have been successfully met and we can get good transcribed text summarization and translation

Sample Example of Summarization and Translation

```
[76]  1 #Text Summary
      2 print(f"------------------------------Summary Of Transcribed Text In {source_language}----------------------------------------")
      3 result[0]
      4
```

```
------------------------------Summary Of Transcribed Text In German-----------------------------------------------
'Buch Adaptationen sind sehr seltsame Wurzeln. In meiner Meinung ist der einzige Weg, einen Film auf dem Grunde eines Buches zu genie
zu-Film-Adaptationen ist Guide.Der Film war nicht R.K. Narayans Guide, sondern die Version von Vijay Anand. Der Film verändert ein Bu
ilm. Der Autor Kalki Krishnamurthy wäre sehr glücklich.Das Buch wurde geschlachtet und die Seiten und Bilder flüchten von einem Stück
n. Nehmt den Charakter der Boatwoman, Pungarelli, gespielt von Aishwarya Lakshmi. Kalki introduziert sie während des Sonnenspiels als
hat mir nicht gut gelungen, aber hier funktioniert diese hochwertige, surreale Auffassung wunderschön. Die spielfreie Szene, in der K
tensive Konversation zwischen Trisha und Vikram. Die tragischen tränen, die in Aishwarya Rai Bachchans Auge wachsen.Die flirtausende
i, die sedu...'  time: 7.22 ms (started: 2023-11-15 18:23:25 +00:00)
```

Transcribed Text Translation

```
[79]  1 #Text Translation
      2 print(f"------------------------------Transcribed Text Translation In {target_language}----------------------------------------")
      3 result[1]  #Kindly expand the output by clicking the '...'
```

```
------------------------------Transcribed Text Translation In English-----------------------------------------------
'Hello and welcome to Galata Plus. In this episode we are talking about Maniratham's PS1, Ponin Selvan 1, which is based on Kalki Krishnamu
tars Karthi, Vikram, Jaim Ravi, Trisha, Aishwarya Rai, Jairam and a whole team of others. Book adaptations are very strange roots and in my
njoy a movie on the bottom of a book is to forget the book. Of course, this is impossible because the book is already in your hip and the
emory. But what I mean by the word is just that, forget the fact that you will get the exact book. One of my favorite book-to-film adaptati
e of the greatest hits of Dev Anand, with great S.D. Bermann songs, a great and beautiful Vahida Rehman and a superb director of Vijay Anan
has the original book.The film was not R.K. Narayan's guide, but the version of R.K. Narayan's guide by film maker Vijay Anand. I remember
```