Bangla Chatbot

CSE 425 Project

Group Members:

SHAHRIN SHAFIQ SIMRAN	20101358
MD RASHID SHAHRIAR KHAN	21201585
FARDIN RAHMAN	20101072
REZUANA IMTIAZ UPOMA	19101130

1. PyTorch :

Torch library is used for nlp. We initially used this library for our model but we encountered an assertion error that CUDA was not enabled. We tried to install and rerun it but still the issue persisted. So we changed to tensorflow.

2. Bert model :

When utilized as a language classifier, BERT excels in categorizing and understanding text. However, its proficiency in generating responses is limited, exhibiting challenges in producing coherent and contextually relevant replies.

I. **BertTokenizer** and **BertForSequenceClassification**:

For conversational response generation, this setup is not inherently designed. The responses would typically be predefined based on the classification input.

Errors Faced :

1. The error **"ValueError: Unrecognized configuration <class 'transformers.model.bert.configuration_bert.BertConfig ' >** for this kind of **AutoModel: TFAutoModelForSeq2SeqLM**' indicates that the model we are trying to loads, '**bert-base-multilingual-cased**', is not compatible with the 'TFAutoModelForSeq2SeqLM' class.
2. The error we have encountered **'ValueError: Couldn't instantiate the backend tokenizer to a fast one'**, is indicating that the tokenizer for the model we are trying to use(such as **'google/mt5-small**') requires the **'sentencepiece**' library, which is not installed in our environment.

3. Transformer :

We encountered some issues in the project. We utilized the auto tokenizer from the Transformer library, but it resulted in an error due to the tokenizer's incompatibility with our dataset .

4. TensorFlow Version Error :

We encountered an attribute error in TensorFlow while using a pre-trained BERT model. The error is related to the version compatibility, and it occurred in **TensorFlow version 1.15**. To address the compatibility issue, we upgraded the **TensorFlow version to 2.15.0**. However, this version proved to be incompatible with the BERT model we were using. As a workaround, we made the decision to exclude the BERT model from our implementation

5. Seq2Seq Model :

- **Data Preparation Issue:**

The initial code for separating questions and answers from the dataset failed to work correctly. Questions and answers were mixed up in the lists.

- **Model Training Issue:**

During model training, there were issues with the loss and accuracy metrics at different epochs, indicating possible training problems.

The training loss and accuracy values provided at different epochs raised concerns about model performance.

- **Model Inference Issue:**

During model inference, there was an issue with generating sensible responses. The generated responses were often incomplete or nonsensical.

- **ReduceLROnPlateau Usage:**

The ReduceLROnPlateau callback does not work properly.

- **Mismatch in LSTM State Size:**

You faced a ValueError due to incompatible state sizes between the bidirectional LSTM encoder (512 units) and the decoder LSTM (256 units).

- **Attention Layer Shape Mismatch:**

The Attention layer threw a ValueError because of mismatched tensor shapes between the decoder's output and the encoder's output.

- **Tokenizer Issues:**

We encountered a KeyError for the newline character ('\n') not found in the tokenizer's word index, indicating that special characters were possibly not included in the tokenizer.

- **Model Input and Output Shapes Mismatch:**

The model architecture during inference had to be consistent with training, focusing on the shapes of inputs and outputs for the LSTM layers.