Information Retrieval and Extraction

Interim submission

Text Style Transfer for paraphrasing

Team : iRetrievers (Team No. 7)

- Pranshu Nema : 2022202029
- Advait Shrivastava : 2022201069
- Ayush Lakshakar : 2022201051

About Datasets

We have used the following dataset in the various stages of our project.

1. ***Formality transfer dataset (GYAFC)***.

    - About: Grammarly's Yahoo Answers Formality Corpus (GYAFC) is a large dataset of informal and formal sentence pairs that is used to train and evaluate language models for formality style transfer. The dataset contains over 100,000 sentence pairs and is the largest dataset of its kind.
    - Use: We might use this if we go for style transfer implementation in the formality class. This will be used to train an inverse paraphraser for formality style transfer (Formal -> Informal or vice versa).

2. ***The Shakespeare dataset & Shakespeare imitation dataset.***

    - About: These data sets contain various works of Shakespeare, and his lite.
    - Use: We Shakespeare imitation dataset (Nonparallel) style transfer we use this in our proposed model for training the inverse paraphraser. We have used Shakespeare dataset(parallel) for the baseline implementation (Fine tuning dataset for the parallel corpus).

3. ***ParaNMT-50M***

    - About: - This dataset is generated as an output of machine translation. We have used this because backtranslation (English -> Czech -> English) introduces some noise in the output which makes it differ from the input in lexical and syntactic choice. We do not use this entire data set as it is. This dataset is aggressively filtered by the authors down to around 75K pairs.
    - Use: - This dataset has been used to train our diverse paraphrasers which are used to make any styled text style neutral.

Note -

- GYAFC and filtered version of ParaNMT-50M are not freely available and were requested from the authors.

Problems in existing approaches of text style transfer

Modern style transfer defines the task of style transfer as modifying the style of a given sentence without appreciably changing the context or semantics of the sentence. So, text style transfer can be assumed to be paraphrasing in the target style. Also, style can be loosely defined as the common patterns of lexical choice and syntactic formulation that are distinct from the subject matter of the sentence, but many existing models fail to distinguish between this.

Also, many Existing methodologies warp the input meaning through attribute transfer technique, but it may lead to change in semantic properties like sentiment. In real world setting, attribute transfer technique is not recommended because preserving the semantic meaning in critical while transferring the style. Semantic preservation is critical for data augmentation, text simplification, writing assistance and many other real-world applications.

Proposed Methodology

➢ *Baseline Models: -*

- ***Zero shot approach -*** In this approach, we attempted to convert an input text from one style to another without giving any examples or training data for that target style. This approach is further explained under section [**Current Work Done].**
- ***Using LSTM to encode and decode –*** This approach considers style transfer as Neural machine translation task. This approach is further explained under section [**Current Work Done].**
- ***Fine Tuning of GPT2 –*** We have finetuned the GPT2 to produce Shakespeare styled text from normal English text. This approach is further explained under section [**Current Work Done].**

➢ *Final Proposed Model: -*

- ***Reformulate Style Transfer as paraphrase generation***

As mentioned above, existing models do not preserve capture the semantic meaning completely, the task can be modelled to ***controlled paraphrase generation*** model. Proposed methodology has 3 stages -

1. Creation of pseudo parallel data by feeding sentences from different styles using ***diverse paraphraser*** model.

2. Train ***Inverse Paraphraser*** model that converts the reformulated sentences back to their source style.
3. Use Inverse paraphraser of target style to perform final style transfer task.

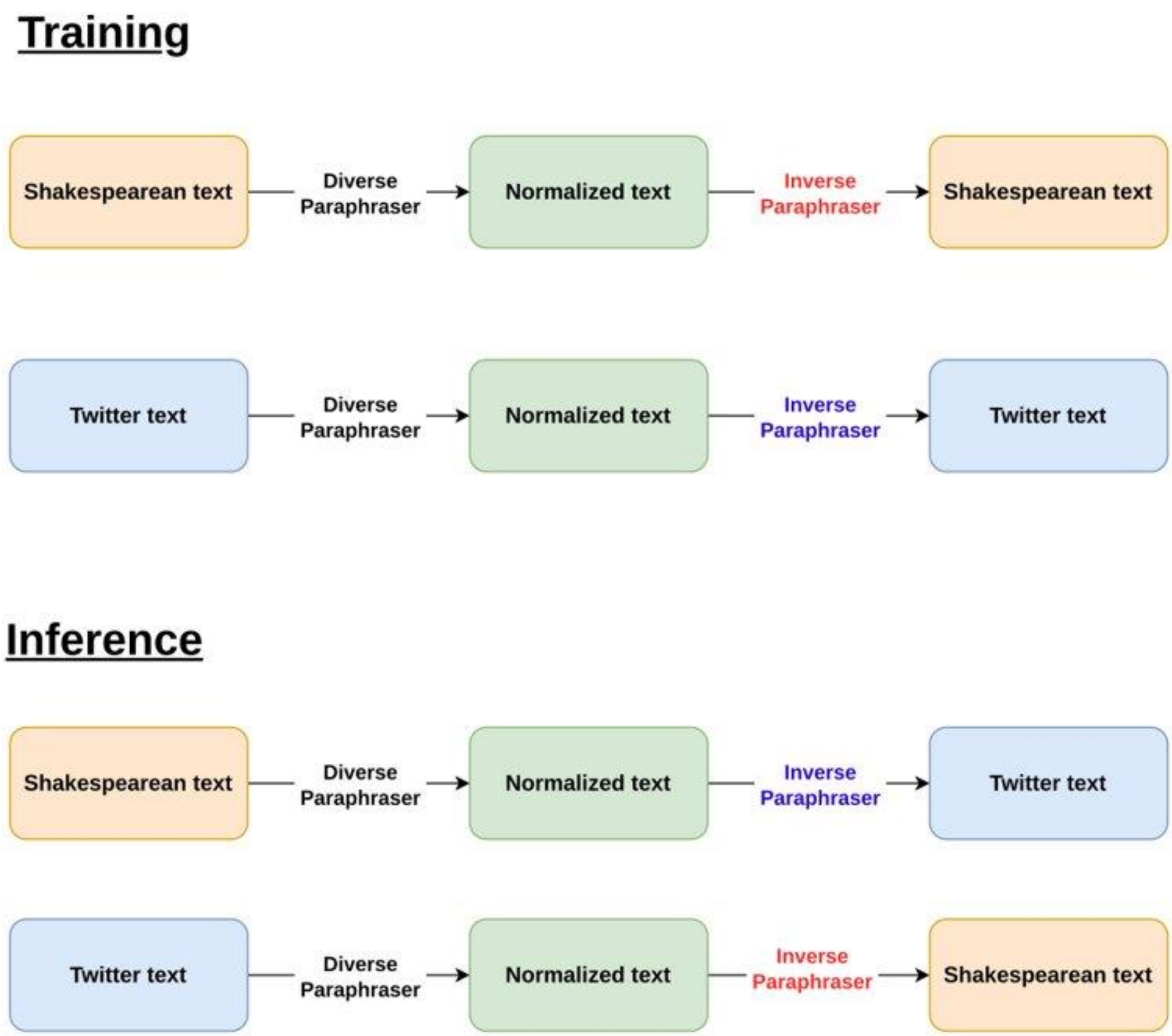

Figure 1 - Proposed Model

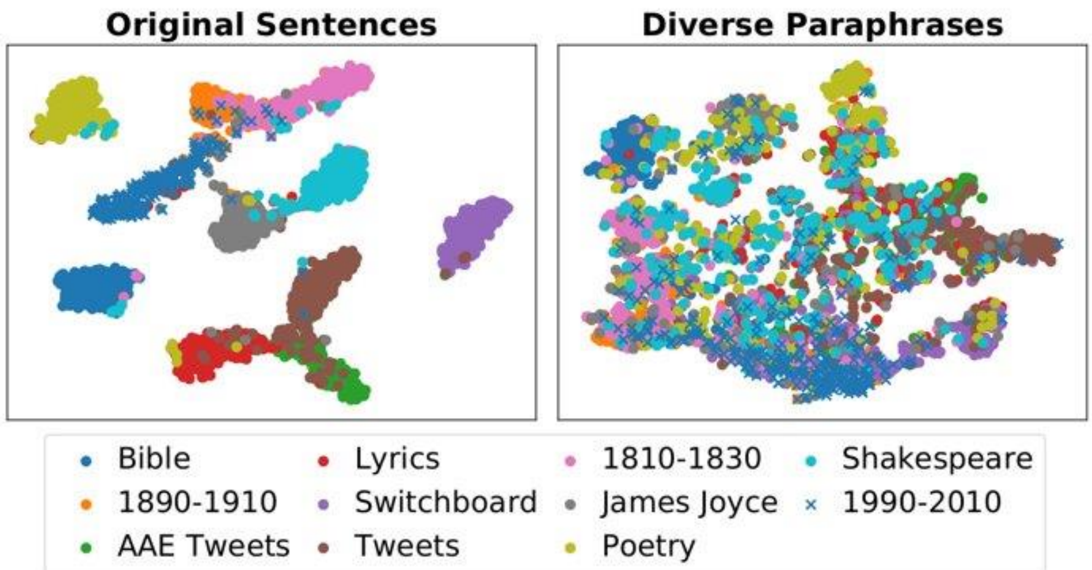

Figure 2 - Diverse paraphrasing normalizes sentences by removing stylistic identifiers.

Current Work Done

➢ Zero shot approach

- In this we tried to transform a given text from one style into another without having access to any specific examples or training data for that style.
- Zero-shot style transfer involves changing the writing style of a text in a completely unsupervised manner, relying solely on the model's pre-trained knowledge and the style description provided.
- Steps we have performed in zero-shot text style paraphrasing
  - Input Text: We start with an input text written in a particular style. This could be formal, informal, Shakespearean, twitter style, or any other style.
  - Style Description: In this we provided a textual description of the desired target style, often in the form of a prompt. For example, "Please rewrite the text in a Shakespearean style."

- Model Inference: Using a pre-trained language model, such as GPT-2 or bart, the model tries to create a text that matches the specific style we want, as described in the style instructions.

- The results generated were not accurate and we were getting unexpected results.

```
Paraphrase this normal english style to Shakespearean style : Hi how are you.
Paraphrased Result : Hi, how's your day.
```

Figure 3 – Zero-shot Result

Finetuning approach

- In this approach we use a parallel dataset to train the GPT2 for style transfer. GPT 2 is a word generational model so to tackle this we must train it to generate Shakespeare, or any style transferred text once fed with the original styled text which can either be no styled or some styled.
- For achieving the above we took the parallel Shakespeare dataset in which we concatenated the normal English sentence and the corresponding Shakespearean sentence for the same.
- Ex: -

Original (English) - *I have half a mind to hit you before you speak again.*

Shakespearian - *I have a mind to strike thee ere thou speak'st*

Transformed - *<s>I have half a mind to hit you before you speak again .</s> >>>>>> <p>I have a mind to strike thee ere thou speak'st .</p>*

- We finetuned our model on corpus with around 18K pairs for 3 epochs and got following validation and train loss.

Epoch	Training Loss	Validation Loss
1	2.267500	2.008694
2	2.088900	1.988575
3	1.941100	1.991719

Figure 4 - Training v/s Validation Loss

```python
# Test the model with a prompt
prompt_text = "<s> Hi How are you .</s> >>>>>> <p>"
responses = generate_response(prompt_text, model, tokenizer)

for response in responses:
    print(response)
```

Figure 5 - Input in simple English/ Modern English

```
<s> Hi How are you.</s> >>>>>> <p>How now, sir?</p>
```

Figure 6 - Output Obtained in Shakespearean style

- The above model provides a decent conversion from normal style to Shakespearian style and serves as a good baseline model for comparison with our proposed scheme. Although it has been observed for some examples that it completely loses the semantic meaning and hallucinates some irrelevant content too.

```
<s> I am so sad today.</s> >>>>>> <p>I am sad now., Come, come, I'll weep.<p>"
<d>You're not going to be able to sleep tonight?</
```

Figure 7 - Hallucinations in generation

➢ *Using LSTM to encode and decode*

- Here, style transfer task is assumed as Neural machine translation task. Shakesperean sentences are encoded to a contextualized representation. The encoded representation is then passed to decoder that has been trained on English sentences. The decoder will then generate the English sentences, corresponding to passed Shakespearean sentence.
- For designing the Encoder, we have used Bi-LSTM and for the decoder purpose, we have used simple LSTM. The same idea will be then re-formulated in final model i.e. style transfer using controlled paraphrase generation.
- Encoder
- The loss curves for encoder and decoder are shown below -

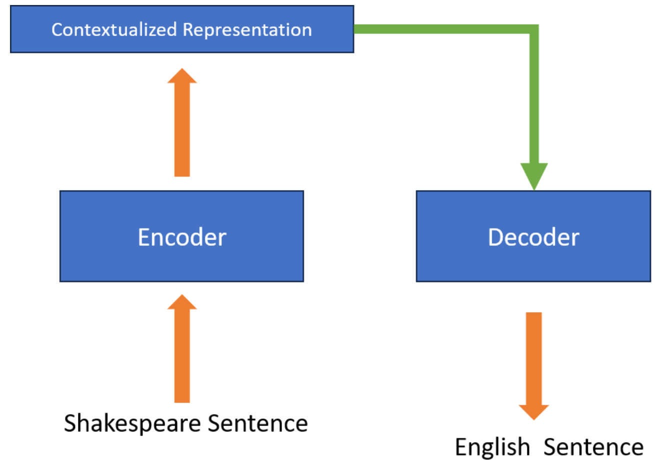

Figure 8 - Encode-Decoder based Architecture

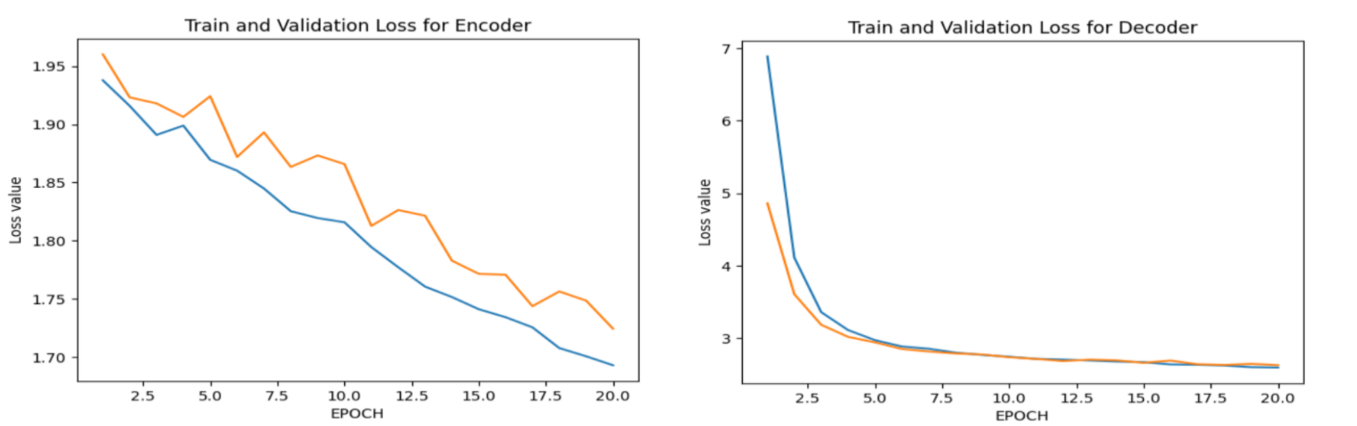

Figure 9 – Encode-Decoder Loss by using LSTM and Bi-LSTM

Assumptions (if any)

- As of now, we have not made any assumptions in the project.

Challenges

The task of altering the style of a given text while preserving its content comes with several challenges. Some of the challenges that we have faced are: -

- ***Lack of data:***
  - Style transfer models typically require parallel datasets containing content in one style and its corresponding version in another style. It was a bit challenging to gather data, compounded by the limited size of the dataset, as constructing parallel datasets can be costly.
  - The dataset was not freely available. We had to email several authors multiple times to obtain the dataset they used in their research papers. We couldn't start our implementation until we obtained the dataset.
- ***Resource Limitations***:
  - Though google collab provide free access to GPU resources but it's just for a limited time. Fine tuning large language models typically requires high computational power and the limited resources in Google Collab were not sufficient for time-consuming fine-tuning tasks.
  - Many times, our sessions were used to time out due to resource limitations of google collab, which required us to restart our training process.
  - We requested access to an ADA cluster to facilitate training our model without resource limitations, but we have not been assigned one yet.
- ***Evaluation Metrics -***
  - Assessing the quality of style transfer models is difficult because there is no definitive metric for measuring how well they have achieved the desired style transfer while preserving content. We must employ a human evaluation method to determine whether the model is functioning correctly, a process that requires a significant amount of effort.

Comparison with deadlines

- As mentioned in the project outline, by 15th September 2023, we were expected to complete Dataset, Baseline exploration and selection. By 31st September 2023, we were expected to complete baseline understanding and exploration.
- Till now as mentioned in **[Current Work Done]** section; we are done with baseline exploration and their implementation. However, we have not started with our final model implementation because we recently got access to the datasets from the authors. Also, due to the unavailability of ADA cluster, we have to work with Google Colab and the fine tuning of GPT2 was taking a lot of time which led to session expiration multiple times. As we now have access to the datasets. We can proceed with our final model implementation i.e. achieving text style transfer via paraphrase generation.

Future Planning

- As we are done with 3 baselines for our proposed model, we have also received the datasets which we requested from the authors.
- We will now be starting with the implementation of proposed model which is '**unsupervised style transfer as paraphrase generation**'.
- In this week we will start with the training of our diverse paraphraser. We plan to use ParaNMT-50M for training it. We intend to finish our diverse paraphraser by the end of this week. By next week, our aim is to start with training inverse paraphrasers. We have planned to evaluate our proposed model on various evaluation metrics and in testing various hyperparameters in the first two weeks of November.