Contextual Text Style Transfer

P. Champion, J.-B. Gaya and M. Herrera-Martinez

Sorbonne Université

Motivation & Background

Text style transfer consists in the translation of a sentence into a desired style (e.g. for sentiment manipulation and formalized writing).

Previous work relied mostly on parallel corpora with a sentence-to-sentence learning framework [2]. This approach requires however a parallel corpus of sentences in both styles, compared on a one-to-one basis. In practice however, such an information may not be available, because sentences are embedded within a paragraph -the dataset is then referred to as non-parallel.

In the studied paper, the approach is to perform the style traner task while maintaining the sentence coherent with its surrounding context [1]. Originally, it consists in a semi-supervised paradigm and the parallel model is implemented and extended with its findings displayed on this poster.

Dataset & Methodology

Dataset

The dataset we created consists in the concatenated corpus of Shakespeare's plays, in its original version and in modern English. Plays and their translations are written in verses providing a parallel dataset. 21079 verses were considered overall.

For each verse x, the associated context ctx_x is defined as the verse immediately before and after the considered verse - except when the previous or the next verse belongs to an other play or does not exist: in that case, the two next or previous verses are considered.

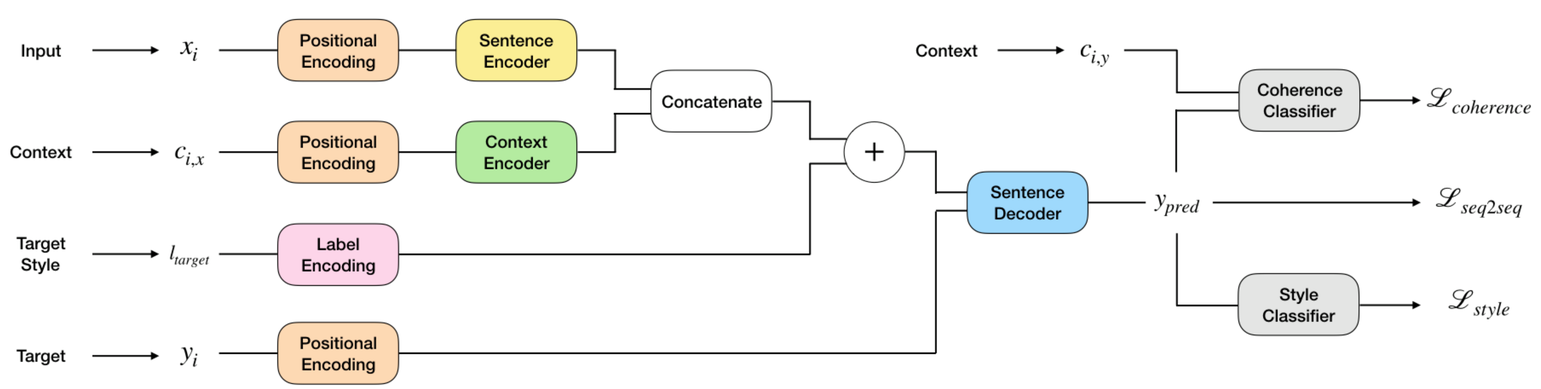
A lowercase method was applied to the whole dataset and some of the punctuation was removed to reduce the size of the dictionary. After processing, the dictionary size equals to 17513.

Methodology

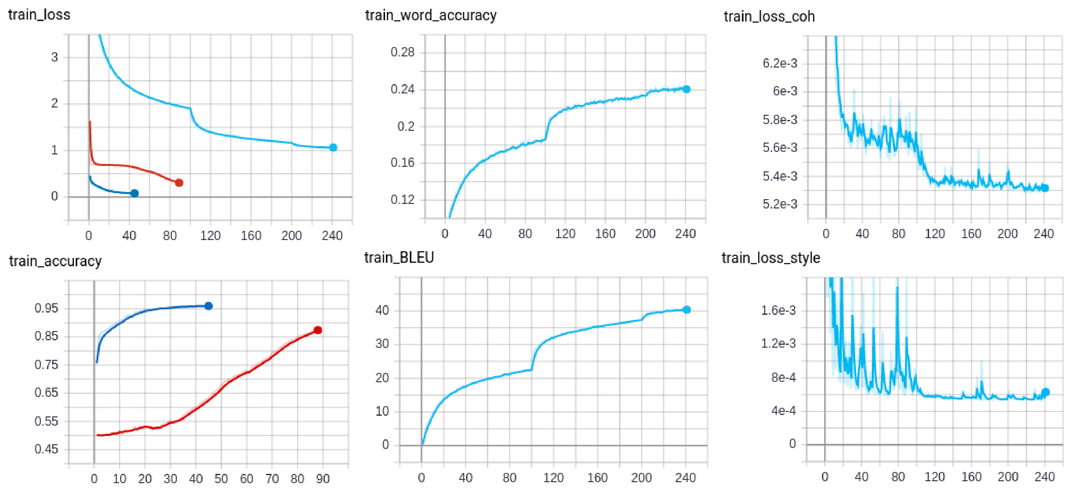
In addition to the **style transfer** task, the main idea of the considered paper is to leverage both a **style classification** (in our case learn whether a sentence is written in modern or Shakespearean English) and a **coherence classification** (evaluate the closeness between the sentence and its context). Since these tasks differ in complexity, a sequential strategy was followed to train the model:

- **Step 1:** Train the **Style classifier** with a DistilBERT pre-trained architecture and a custom embedding of dimension 768, for it to be compatible with the DistilBERT model used in the style and coherence classifier [3]. **Hyperparameters** : batch size: 64 | learning rate: 1e-5 | optimizer: Adam | Epochs: 50 | Accuracy reached : 95%.
- **Step 2:** Use the Style classifier as a pre-trained model, freeze the custom embedding and train the **Coherence classifier** while fine-tuning. **Hyperparameters** : batch size: 64 | learning rate: 6e-5 (with warmup strategy) | optimizer: Adam | Epochs: 90 | Accuracy reached : 86%.
- **Step 3:** Train the **Style Transfer** model while freezing both the Style the Coherence Classifiers. The encoders and decoder consist in a one-layer, 4-headed transformers with a feedforward dimension of 1024. We trained it 3 times, which explains the drops in the results **Hyperparameters** : batch size: 32 | learning rate: 5e-4 -> 1e-5 | optimizer: Adam | Epochs: 240 | BLEU score : 40.70%.

Architecture



Results



Left : train losses and accuracy of the style classifier, coherence classifier, global model. Middle : sentence accuracy and BLEU score of the global model. Right : the impact of the style and coherence losses during the training of the global model.

Input	Input style	Target	Translation
<SOS> is it possible that any crime could be so valuable ? <EOS>	Modern	<SOS> is it possible that any villainy should be so dear ? <EOS>	<SOS> is it possible that any villainy should be unworthy of my wife ? <EOS>
<SOS> good my mouse of virtue , answer me . <EOS>	Shakespearian	<SOS> please answer , my good little student . <EOS>	<SOS> good my good friend , answer me . <EOS>
<SOS> your husband being troubled with a shrew measures my husbands sorrow by his woe . <EOS> .	Shakespearian	<SOS> your husband , being saddled with a shrew , projects his own suffering onto my husband . <EOS>	<SOS> your husband being saddled a shrew , projects his own suffering onto my husband <EOS>

Conclusion

Key take-aways

- The architecture is a mix of pre-existing structures used in other papers.
- The use of two additional losses does seem to contribute to the overall performance.
- The translated phrases are close to the target sentences semantically.

Future works

- Train the embedding during the main task.
- Consider the use of other kind of label encoding.
- Implement the non-parallel architecture while considering context.

References

[1] Yu Cheng et al. *Contextual Text Style Transfer*. 2020. URL: `https://openreview.net/forum?id=HkeJzANFwS`.

[2] Harsh Jhamtani et al. *Shakespearizing Modern Language Using Copy-Enriched Sequence-to-Sequence Models*. 2017. arXiv: `1707.01161 [cs.CL]`.

[3] Victor Sanh et al. *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. 2019. arXiv: `1910.01108 [cs.CL]`.