

# 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. It encompasses various applications, e.g. sentiment manipulation and formalized writing. Previous work relied mostly on parallel corpora with a sentence-to-sentence learning framework [bahdanau2014neural],[jhamtani2017shakespearizing]. 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, an approach to perform the style transfer task while maintaining the sentence coherent with its surrounding context is discussed. This approach is mostly targeted at non-parallel data but may be extended to parallel data.

This poster summarizes the work conducted based on this paper. The section ?? introduces the key concepts and modules used to build the model: in this section we notably elaborate on previous work and underline our contributions. The section summarizes the results of our approach.

## Model

### Dataset

The considered dataset consist in the concatenated corpus of the Shakespeare’s plays, in its original version and in modern English. Plays are written in verses, as well translations, enabling 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, it is ignored.

### Architecture

#### Embedding

An uppercase method was applied to the whole dataset and after some further post-processing[lequel]. With this processing, the dictionary size equal to 17513. An embedding dimension of 768 was considered, for the embedding to be compatible with the BERT model used in the coherence classifier [ref].

The embedding was trained jointly with the Style Classifier (see subsection on the Style Classifier)

#### Encoder

#### Decoder

#### Coherence Classifier

T

#### Style Classifier

The Style classifier takes as inputs two sentences  $x_1, x_2$  and predicts their respectives styles  $(\hat{l}_1, \hat{l}_2)$ . It is evaluated with a binary cross entropy loss against their actual style  $(l_1, l_2)$ .

For simplification purposes, the embedding layer was prepended to the style classifier and both modules were trained together. The training took place over the whole dataset and plateau with an accuracy of xx % (vs. 86.78 % in the reference paper).

## Results

## Remarks

## References