Style Transformer

Introduction:

Text style transfer is the task of changing the stylistic properties of the text while retaining the style-independent content within the context.
Neural networks have become the dominant method in text style transfer.
Most previous methods formulate the style transfer problem into the "encoder-decoder" framework.
The encoder maps the text into a style-independent latent representation, and the decoder generates a new text with the same content but a different style from the disentangled latent representation plus a style variable.

· They introduce a novel training algorithm that makes no assumptions about the disentangled latent representations of the input sentences, and the model can employ attention mechanisms to improve its performance further
· Experimental results on two text style transfer datasets have shown that our model achieved a competitive or better performance compared to previous state-of-the-art approaches
· The back translation technique developed by Lample et al. (2019) can be adapted to the training process of Style Transformer

They use the highly polar movie reviews Maas et al. (2011) provided to get a high-quality dataset. Based on this dataset, they construct a highly polar sentence-level style transfer dataset by the following steps: 1) fine-tune a BERT (Devlin et al., 2018) classifier on the original training set, which achieves 95% accuracy on the test set, 2) split each review in the original dataset into several sentences; 3) filter out sentences with confidence threshold below 0.9 by our fine-tuned BERT classifier; 4) remove sentences with uncommon words

Objectives

To tackle the style transfer problem the authors defined above, the goal is to learn a mapping function $f\theta(x, s)$ where x is a natural language sentence, and s is a style control variable.

A goal-transferred sentence should be fluent and content-complete with a target style. Following previous works to evaluate the performance of the different models, the authors compared three different dimensions of generated samples: 1) Style control, 2) Content preservation, and 3) Fluency.

Results

Style Control: The authors automatically measure style control by evaluating the target sentiment accuracy of transferred sentences.

A higher BLEU score indicates that the transferred sentence can better preserve content by retaining more words from the source sentence.
The models achieve overall competitive performance and better content preservation in the two datasets compared to previous approaches.
The authors' conditional model can better control style than the multi-class model.
Both models can generate sentences with relatively low perplexity.
The authors choose two of the most well-performed models according to the automatic evaluation results as competitors: DeleteAndRetrieve (DAR) (Li et al., 2018) and

Conclusion

The authors proposed the Style Transformer with a novel training algorithm for text style transfer tasks.
Experimental results on two text-style transfer datasets have shown that the model achieved a competitive or better performance than previous state-of-the-art approaches.
The model can get better content preservation on both datasets because the proposed approach does not assume a disentangled latent representation for manipulating the sentence style.
The authors plan to adapt the Style Transformer to the multiple-attribute setting like Lample et al. (2019).
The back translation technique developed by Lample et al. (2019) can be adapted to the training process of Style Transformer.

Controllable Unsupervised Text Attribute Transfer:

Text attribute transfer is text editing to alter specific attributes, such as sentiment, style, and tense.

· The dominant methods of unsupervised text attribute transfer are to separately model attribute and content representations, such as using multiple attribute-specific decoders or combining the content representations with different attribute representations to decode texts with target attributes in an adversarial or non-adversarial way
· Because they try to disentangle attribute and attribute-independent content, this may undermine the integrity and result in poor readability of the
· The authors present a controllable unsupervised text attribute transfer framework, which can edit the entangled latent representation instead of modeling attribute and content separately.
· This is the first to control the transfer degree freely and perform sentiment transfer over multiple aspects simultaneously.

· The authors find that there may be some failure cases, such as learning some attribute-independent data bias or just adding phrases that match the target attribute but are useless.