Demo: Literary Style Transfer with Content Preservation

Katy Ilonka Gero, Chris Kedzie, Lydia B. Chilton
Columbia University
{katy,kedzie,chilton}@cs.columbia.edu

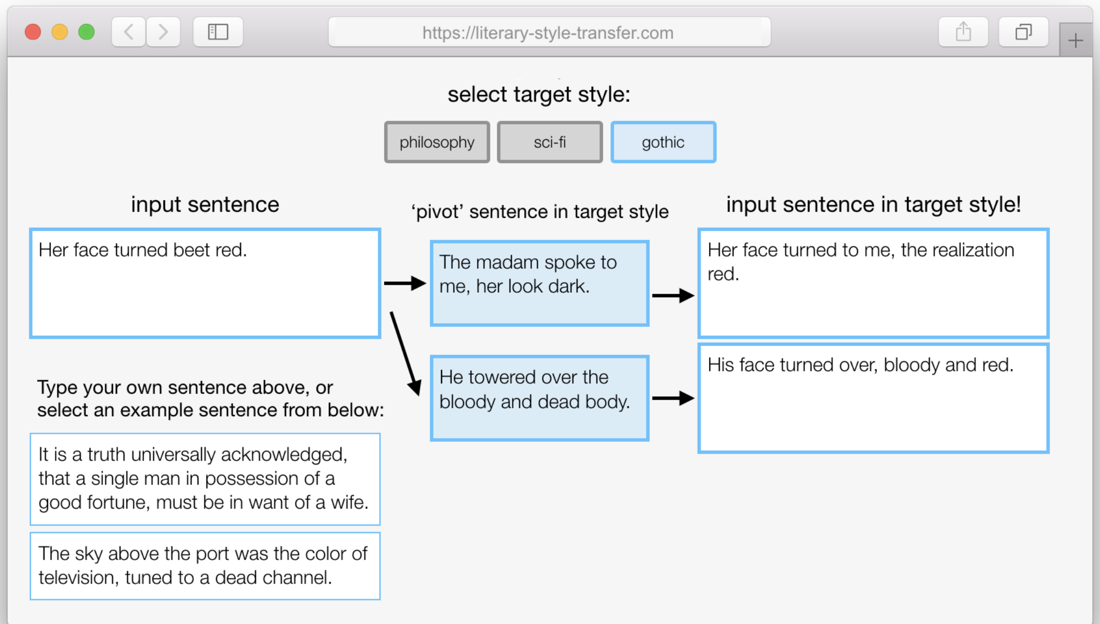

Figure 1: Mock-up of demo. Users can input their own sentence and 'transfer' it to an alternative literary style.

ABSTRACT

Successful style transfer in images has led to fruitful human-AI collaborations, as both novice and expert imaginations are sparked by high-fidelity and intuitive image outputs. But style transfer for text is still an open challenge. Drawing on computational studies of literature, we present a neural encoder-decoder model that transfers sentences from one literary style to another. We model literary style as a suite of low-level linguistic controls, such as frequency of pronouns, prepositions, and subordinate clause constructions. We perform style transfer by keeping the content words fixed while adjusting the controls to be indicative of another style. In this demo, users are able to type in their own sentences and transfer them into one of three literary styles: philosophical, gothic horror, or science fiction. This demo is presented as a web interface, and allows users to transfer any sentence or paragraph into a desired style.

KEYWORDS

natural language processing, human-computer interaction, writing support, style transfer, computational literature

ACM Reference Format:

Katy Ilonka Gero, Chris Kedzie, Lydia B. Chilton. 2020. Demo: Literary Style Transfer with Content Preservation. In *IUI '20 Workshops, March 17, 2020, Cagliari, Italy*. ACM, New York, NY, USA, 2 pages.

1 INTRODUCTION

The success of style transfer in image processing has sparked the imagination of novices and experts alike. From the outputs of professional artists like Helena Sarin to popular demos like Google Deep Dream, these tools have fueled human-AI collaboration. Yet we live in a deeply literate society, and write far more often than we produce images, so what about manipulations of text?

All text has style, whether it be formal or informal, polite or aggressive, colloquial, persuasive, or even robotic. Despite the success of style transfer in image processing [3, 4], there has been limited progress in the text domain, where disentangling style from content is particularly difficult.

In this work, we turn to literary style as a test-bed for style transfer, and build on work from literature scholars using computational techniques for analysis. In particular we draw on stylometry: the use of surface level features, often counts of function words, to discriminate between literary styles [6, 7].

We present a controllable neural encoder-decoder model in which these surface-level linguistic features are modelled explicitly as decoder feature embeddings. This model takes in an input sentence and target style and outputs a new sentence in the target style. Our demo embeds this model into a web interface that allows users to perform style transfer on arbitrary sentences as well as play with and control various aspects of the style transfer process.

2 MODEL DESIGN, TRAINING & EVALUATION

In this section we give an overview of the model; full implementation and evaluation details can be found in [5]. We implement our feature-controlled language model using a neural encoder-decoder with attention [1], with 2-layer unidirectional gated recurrent units (GRUs) for the encoder and decoder [2]. This model takes in a sequence of words and a set of feature controls which represent the target style. It outputs a new sequence of words – the original sentence in a new style.

The feature controls are a set of counts of stylistically relevant sentence features. There are 17 features in total, and include features such as the number of conjunctions, negation words, and helper verbs. These features represent the desired style of the output sentence, and can be calculated for any sentence.

In training, the model reconstructs a sentence using only the content words of the input sentence and the input sentence's own feature controls. We use a corpus of literary texts including philosophy books, gothic horror novels, and science fiction short stories.

For example, in training on the sentence 'The vampires fly in space.', the model would have an input like:

content words: vampires fly space
feature controls: [conj: 0, article: 1, preposition: 1, ...]

And the desired output would be the original sentence:

desired output: The vampires fly in space.

With a trained model, we transfer arbitrary content words to a new style without parallel data by setting the feature controls to be indicative of the target style. The new feature controls can be found in a variety of ways; a baseline method may be to select controls that represent the average features of a given corpus. However, the discrete nature of this task, and the variety of ways style is surfaced across the thousands or millions of sentences in a corpus, makes averaging less useful that may otherwise be supposed.

Instead we use a 'pivot sentence', which is a sentence in the target style with a similar length and parts of speech. We calculate the desired feature controls using this sentence. This method also allows us to increase the diversity of transferred sentences by changing the pivot sentence selected, and allows an easy-to-understand entry point for a user interested in exploring abilities of our model.

Here's an example of how style transfer would be performed on 'Her face turned beet red.' with the target style 'gothic horror':

content words: face turned beet red
pivot sentence: the madam had become cold, her look dark.
feature controls: [article: 1, pronoun: 1, punct: 2, ...]
output sentence: her face had turned to me, the realization red.

In an automatic evaluation, our transferred sentences are classified as their target style 84% of the time. We also find that our model produces highly diverse and stylistically distinctive outputs compared to a baseline method. Full details can be found in [5].

3 USER INTERACTION

We embed this model into a web interface, as seen in Figure 1. This interface allows users to type in whatever text they would like, or select from a large set of example sentences from famous literary texts. The user must also select a desired target style. The system then automatically selects several pivot sentences, and uses these pivot sentences to generate a new sentence in the target style. The user can then experiment with transferring to different styles or using different input texts. Additionally, the pivot sentences expose the user to part of how the model works. Users can directly modify the pivot sentences and re-generate the transfer sentence. This both gives users some insight into how the style transfer is occurring, as well as giving them more control over the final sentence.

Although we have discussed single sentences, this system is able to transfer any length of text by splitting it into sentences. Therefore users can actually input paragraphs of text and see these entire paragraphs transferred into a new style.

4 LIMITATIONS AND FUTURE WORK

By encouraging the model to preserve the content words, we do not take advantage of replacing content words with synonyms or more drastic paraphrasing that preserves meaning. Additionally, we found it difficult to control larger-scale syntactic structures, like rearranging clauses. In the future, we could address these shortcomings with noising methods and specialized training objectives.

This demo allows users to play with a model for the computational style transfer of text. Style is a key element of writing, and future human-AI writing systems will likely find style transfer, or at least careful style consideration, to be an important functionality. These writing systems may help users revise existing writing to better match the desired style, or may use style transfer as a way to challenge users to consider their stylistic choices.

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
[2] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP).* 1724–1734.
[3] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576* (2015).
[4] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2016. Image Style Transfer Using Convolutional Neural Networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*
[5] Katy Gero, Chris Kedzie, Jonathan Reeve, and Lydia Chilton. 2019. Low-Level Linguistic Controls for Style Transfer and Content Preservation. *arXiv preprint arXiv:1911.03385* (2019).
[6] Frederick Mosteller and David L. Wallace. 2007. *Inference and Disputed Authorship: The Federalist.* Center for the Study of Language and Information.
[7] Jonathan Reeve. 2019. On Early Style: a Stylochronometric Critique of Late Style in Literature. *Under Review at Digital Scholarship in the Humanities* (2019).