Unpaired Abstract-to-Conclusion Text Style Transfer using CycleGANs

Haotong Wang
Waseda University
Kitakyushu, Japan
wanghaotong0925@toki.waseda.jp

Yves Lepage
Waseda University
Kitakyushu, Japan
yves.lepage@waseda.jp

Chooi Ling Goh
The University of Kitakyushu
Kitakyushu, Japan
goh@kitakyu-u.ac.jp

***Abstract*—The availability of paired examples greatly facilitates the task of style transfer by allowing the use of supervised learning. However, our scenario does not enjoy such a condition. We focus on style transfer for academic writing, and examine the possibility of performing style transfer between sentences from the abstract and conclusion sections of a scientific article in the Natural Language Processing field, in both directions. We assume a latent correlation between the abstract and conclusion styles, and construct an unpaired data set. We propose the use of a version of `CycleGAN` based on transformers to perform the task. Our approach is shown to realize differences in tense or word usage which are characteristic of the different sections.**

***Index Terms*—Text style transfer, unpaired data, GANs.**

I. Introduction

Writing aid system aims at assisting people in composing a text. The main function is to ensure correct expression. In the case of a researcher writing a scientific article, the language proficiency is essential. Improper style in the different sections of the article is one of the problems that non-English native speakers may be confronted to.

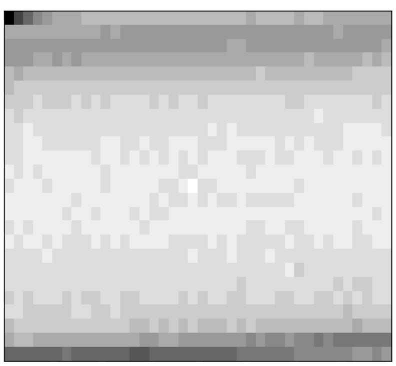

Fig. 1. Chance that a token at a given position in an article appears in the abstract. The darker the higher the chance. Average over 15,000 articles from the ACL-ARC collection. Image copied from [1]

Figure 1 materialises the positions of words in a scientific article, from the introduction to the conclusion, averaged over an entire collection of scientific articles. A darker pixel indicates that the words at that position have a higher chance to appear in the abstract of the same article. The figure shows that words in the abstract are relatively much used in the introduction, then less and less used in the body of the article, and finally very much used in the conclusion. This conforms with human empirical knowledge that an abstract borrows much of its content from the introduction and the conclusion.

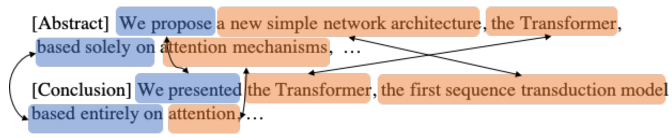

Fig. 2. Similar sentences in the abstract and the conclusion. The sentences are extracted from an article published in NIPS 2018: [2].

Indeed, the abstract and conclusion sections have identical functions in a scientific article: they provide a summary and succinctly describe the content of the research. However, the abstract is a preview while the conclusion is a review. This leads to the use of different grammatical or rhetorical devices. Figure 2 shows a pair of sentences from the same article, the third sentence of the abstract and the first sentence in the conclusion. The sentences have basically the same meaning, but they exhibit differences in tenses (present/past), word usage (propose/presented) and structure (apposition/anteposition).

Our ultimate goal is to help researchers in writing scientific articles and reduce their effort in this process. We thus exploit the result of the previous visualisation and propose to help researchers in writing sentences in a conclusion from sentences in the abstract, or conversely. This amounts to performing style transfer from the conclusion ‘style’ to the abstract ‘style’.

To this end, we build a model which generates texts in both directions, from abstract to conclusion, and from conclusion to abstract, while exchanging the characteristics of the ‘styles’. Two main concerns need to be dealt with:

- Absence of a clear definition of abstract ‘style’ and conclusion ‘style’;
- Absence of directly available data parallel in content for the two ‘styles’ at hand;

Unsupervised text style transfer focuses at controlled language generation which aims to transfer the given style while preserving the original content without parallel data [3]. A major problem is the difficulty in collecting aligned sentences with the same content but opposite styles. Previous works mainly focus on unsupervised learning method to tackle that. It has been applied in many well-defined style tasks, such as sentiment transfer [3], [4], [5], [6], [7], [8] or formality transfer [9], [10]. These styles have distinct characteristics. They belong to two poles in styles, which can be easily identified by human beings. When creating such corpora, people firstly define the notion of opposite style (e.g., positive vs negative opinion), filtering the data so that it can be transferred. As for style transfer between abstract and conclusion sections, because we lack a clear definition of style, we expect to discover the latent style automatically instead of defining it in advance.

II. Contributions of the paper

In this paper, we assume that the different styles in the abstract and conclusion sections of a scientific article can be represented by some latent representations which can be used to transfer one style onto the other one. In order to construct a corpus to train from, we use the ACL Anthology Reference Corpus (ACL-ARC)[1] as our primary data. It contains thousands of high quality scientific articles in the Natural Language Processing (NLP) field. To retrieve and assemble pairs of similar sentences from the abstract and conclusion sections of articles, we use Sentence-BERT [11] to obtain vector representations of sentences and compare sentences by cosine similarity of their vector representations.

A transformer-based Cycle-GAN is then implemented to perform style transfer in both directions: from abstract style to conclusion style, and from conclusion style to abstract style.

Our contributions are summarized as follows:

- We introduce a method to organize a data set, composed of sentences in the field of NLP with their style, i.e., whether they belong to the abstract or conclusion sections.
- We propose a Transformer-based `CycleGAN` model, which makes no assumption about the disentangled latent representations of the input sentences. We train this model on our data set.
- We perform experiments the results of which show that our proposed model generally outperforms other approaches on this same data set.

III. Related work

A. Text style transfer

The classical way of addressing the problem of text style transfer is to rely on a large amount of pairs of examples with the same content, but differing in styles, usually opposite styles. Several data sets exist like the Yelp Review Dataset[2] and Amazon Food Review Dataset [12] for sentiment style transfer, or the GYAFC Dataset for formality style transfer [9], etc. The use of supervised learning is possible on such data sets.

However, for style transfer problems like the one we address, collecting such neatly opposite examples is difficult because it may not be the case that the content and style are neatly separated. Unsupervised or semi-supervised learning is thus the solution to tackle the problem. It relies on the use of unpaired examples.

This vein of research can be split into two approaches. The first approach consists in extracting style from sentences by dissolving content through disentanglement of latent representations. Style transfer then consists in combining the content representation of a sentence with the target style representation. For instance, Hu et al. (2017) [13] propose an approach which combines variational auto-encoders and attribute discriminators to impose explicit independence constraints on attributes; Shen et al. (2017) [3] propose a cross-aligned encoder-decoder architecture aiming to leverage refined alignments of latent representations; Prabhumoye et al. (2018) [5] learn a latent content representation using back translation and adversarial generation to match the output onto the desired style.

The second approach aims to learn and transfer the sentences without explicitly separating content from style. For instance, Xu et al.(2018) [7] propose a cycled reinforcement learning method which is trained on unpaired data through the collaboration of two so-called 'neutralization' and 'emotionalization' modules; Luo et al. (2019) [6] propose a dual reinforcement learning framework to directly transfer the style of the text via a one-step mapping model, without any separation of content and style; Dai et al. (2019) [4] propose a Style Transformer, which makes no assumption about the latent representation of a source sentence and uses an attention mechanism to improve style transfer performance.

The second approach has been shown to be more effective than the first approach: better performance in both style control and content preservation have been achieved without explicit separation of the content and the style. Consequently, our work positions itself in the second approach.

B. Cycle-Consistent Adversarial Network (`CycleGAN`)

Generative Adversarial Networks (GANs) have achieved success in image processing, and afterwards in NLP in semi-supervised and unsupervised learning tasks like text generation and machine translation [14]. The two components of a GAN are a generator and a discriminator: the generator generates constrained candidates which are evaluated by the discriminator.

An elaboration of GANs is Cycle-Consistent Adversarial Network (`CycleGAN`). It has been shown to be effective in image processing for unpaired image-to-image translation [15]. In `CycleGAN`s, a pair of generators are used to separately perform a forward and an inverse mappings; for us, style transfer in both directions. With this, it is possible to find optimal pseudo pairs of objects from unpaired data sets. A `CycleGAN` may thus be the right choice for the problem we are addressing, where no clearly opposite examples can be found to form neatly aligned pairs of sentences which can be used directly in supervised training.

Several applications of `CycleGAN` have shown their capability in better preserving the content when performing trans-

[1]https://acl-arc.comp.nus.edu.sg

[2]https://www.yelp.com/dataset/challenge

fer. While the original proposal is rooted in image processing, an adaptation of `CycleGAN` has been used at the interface of image and texts for text-to-image generation [16]. In the field of text processing, `CycleGAN`s have also been used for style transfer between news and poems [17].

IV. Method

In this section, we explain how we formalize our problem. We first design an algorithm to construct an unpaired data set from open source articles. We then implement style transfer itself using `CycleGAN`s.

A. Problem Definition

Suppose that we have two sets of sentences labelled by the section they belong to:

$$A = \{(X_i, a)\}_{i=1}^{N} \quad \text{and} \quad C = \{(Y_j, c)\}_{j=1}^{M}$$

A stands for the set of sentences in the abstract section and C for the conclusion. X_i and Y_j denote sentences in different sections, a and c correspond to their labels. The goal of our work is: given a sentence X_i with style a, to change it into some sentence X_i' with style label c. In the opposite direction, we also transfer sentence Y_j with style c onto a target sentence Y_j' with style a.

B. Construction of an Unpaired Data Set

In order to extract the differences in style between the abstract and the conclusion sections of scientific articles, we design a method to extract sentences whose content is similar. We use Sentence-BERT [11] to compute sentence representations, i.e., embedding vectors corresponding to sentences.

We retain sentences in the abstract or the conclusion sections which have a corresponding sentence with a similar content. To this end, we proceed as follows. For each sentence in an abstract section, we search for the most similar sentences in the conclusion section of the same article. We retain at most three sentences which have a similarity above a given threshold. If there is no corresponding sentence in the conclusion, we eliminate the sentence from the set of sentences belonging to the abstract style.

Notice that we do not build a data set of paired sentences. We can only build a data set of unpaired examples, because we can only check for similarity. By definition of similarity, we cannot claim that the sentences in a pair differ only by their style; they may also differ in content for a good amount. For this reason, we forget about which sentence in the conclusion corresponds to which sentence in the abstract, and just retain two sets.

C. Overview of the Model

As illustrated in Figure 3, we synchronously train two different GANs to transfer abstract sentences into conclusion sentences and another one for the other direction. For each GAN, there is a generator and a discriminator. The two GANs ensure the independence in both directions of style transfer, while allowing style learning by adversarial training.

In typical GANs [18], the goal of training is to minimize the adversarial loss $L_{adv}\left(G_{A\to C}\left(a\right), c\right)$, which means that sentences generated from the domain A are as close to the target domain C as possible. `CycleGAN` has two adversarial loss functions in both directions. For the mapping from a sentence (X_i, a) in the abstract section onto a sentence (X_i', c) in the conclusion section, we have:

$$L_{adv}\left(G_{A\to C}\left(X_i, a\right), (X_i', c)\right) \tag{1}$$

In the other direction, from conclusion to abstract, we have:

$$L_{adv}\left(G_{C\to A}\left(Y_j, c\right), (Y_j', a)\right) \tag{2}$$

The cycle in the `CycleGAN` enforces consistency by reconstructing a sentence with its original style so as to guarantee preservation of content. The loss function is defined as :

$$\begin{aligned} L_{cyc} = {} & \mathbb{E}_a[\![G_{A\to C}\left(G_{C\to A}\left(Y_j, c\right)\right) - (Y_j, c)]\!] \\ & + \mathbb{E}_c[\![G_{C\to A}\left(G_{A\to C}\left(X_i, a\right)\right) - (X_i, a)]\!] \end{aligned} \tag{3}$$

Before training the whole network, the generators are trained in advance. A generator is fed with sentences as input from the target domain. It is expected to generate the same sentence without any change, i.e., to perform the identity mapping. The identity loss is defined as:

$$\begin{aligned} L_{identity} = {} & \mathbb{E}_a[\![G_{C\to A}\left(X_i, a\right) - (X_i, a)]\!] \\ & + \mathbb{E}_c[\![G_{A\to C}\left(Y_j, c\right) - (Y_j, c)]\!] \end{aligned} \tag{4}$$

With all the previous losses above, the global objective of a `CycleGAN` is given by:

$$\begin{aligned} L = {} & L_{adv}\left(G_{A\to C}\left(X_i, a\right), (X_i', c)\right) \\ & + L_{adv}\left(G_{C\to A}\left(Y_j, c\right), (Y_j', a)\right) \\ & + \lambda_{cyc} L_{cyc} + \lambda_{identity} L_{identity} \end{aligned} \tag{5}$$

where λ_{cyc} and $\lambda_{identity}$ are hyper-parameters which adjust the balance between identity mapping and the cycle loss function.

D. Transformer-based Generator Network

The generator in our model is implemented as a Transformer [2]. Transformers use multi-head self-attention networks. They are point-wise feed-forward networks and can be used both as encoders or decoders. They have been shown to have a good ability to extract meaning in style transfer [4].

The encoder maps an input sequence of symbols $W = \{w_t\}_{t=1}^{N}$ onto a corresponding sequence of continuous representations $Z = \{z_t\}_{t=1}^{N}$. Given Z, the decoder generates an output sequence of symbols $W' = \{w_t'\}_{t=1}^{N}$, one element at a time. It factors this distribution as:

$$p_\theta\left(W' \mid W\right) = \prod_{t=1}^{m} p_\theta\left\{w_t' \mid Z, w_{<t}'\right\} \tag{6}$$

The parameters θ in the network should minimize the global objective function given in Equation (5). At each time step t,

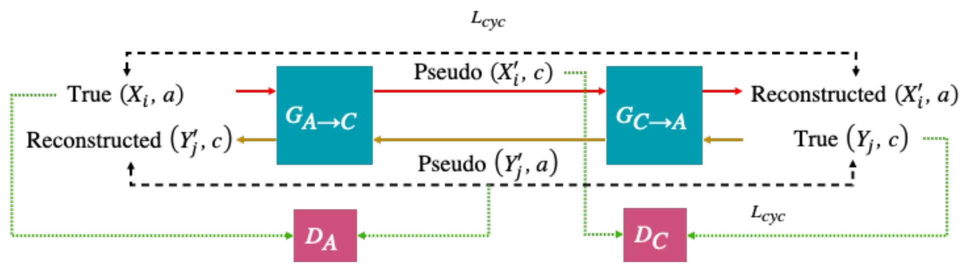

Fig. 3. Our proposed `CycleGAN` architecture. C and A stand for abstract and conclusion. $G_{C\to A}$ stands for the direction of style transfer from conclusion to abstract, and $G_{A\to C}$ is the inverse direction. D is for discriminator and G or transformer-based generator.

TABLE I
PARSING RESULTS FROM ACL-ARC

# of articles used	21,520
# of abstract extracted	19,385
# of conclusion extracted	16,317
# of articles with both sections	15,799

the prediction of the next token is computed by a softmax classifier:

$$p_\theta \left\{ w'_t \mid Z, w'_{<t} \right\} = \text{softmax}\left(o_v\right) \tag{7}$$

where o_v is the last layer output by the decoder network.

E. Discriminator Network

Following [19], our discriminator consists of two layers, a word embedding layer and a fully connected layer. In order to accelerate the training process, bi-gram word features are embedded. Hierarchical softmax is used as the classifier. All this can accurately and efficiently capture the content of a sentence.

V. Experiments

A. Dataset

ACL-ARC is a corpus of scientific articles in the NLP field published by or in association with the Association of Computational Linguistics (ACL). It comprises 21,520 articles from year 1979 to year 2015. From the provided ParsCit structured XML files, we parsed and extracted 19,385 abstract sections and 16,317 conclusion sections. The smaller number of conclusions shown in Table I is explained by articles where the section was absent or was not extracted automatically. The method described in Section IV-B needs articles with both sections. 15,799 articles have both abstract and conclusion sections.

Table II reports some statistics about the data set. 93,214 sentences were extracted from the abstract sections of 15,799 articles, slightly less than the number of sentences in the conclusion sections: 100,787. The average length of a sentence is about 15 words.

We apply the method described in Section IV-B to build our unpaired data set after filtering out any sentence with a length outside of the range 5 to 25 in order to ensure a uniform distribution of data [9]. This results in 44,901 and 48,483 sentences with abstract and conclusion labels, which constitute our unpaired data set. These numbers are relatively balanced. The average length of a sentence decreased to 13 words for both sections. We randomly select 500 sentences for the test set and 500 sentences for the validation set. The remaining is the training set.

B. Evaluation

In the evaluation of text style transfer, three aspects should be taken into consideration.

- Variation in style: it is evaluated by measuring the accuracy on the transferred sentences. For that, we train a different binary classifier based on the transformer to distinguish between the two styles. Its accuracy is xx %.
- Preservation of content: we compute BLEU scores to evaluate how different the transferred sentences generated by each generator are from the input sentences, considered as references. A relatively high BLEU score will indicate that the system preserves content by retaining the words from the source sentence [4].
- Fluency: it is measured by the perplexity of the transferred sentence in a language model. We train a 5-gram language model on the training set using KenLM [20].

C. Experimental Results

1) Impact of Threshold on the Unpaired Data Set: We first determine the best value for the threshold T by measuring its impact on the amount of data and the performance of the system. We test 4 values, ranging from 0.6 to 0.9. The results are shown in Table III.

By definition of the threshold (see Section IV-B), an increase in its value entails a reduction in the number of extracted sentences: roughly 30,000 sentences less for each increment of 0.1. Higher values will lead to a reduction in the diversity available in the data for each style. On the contrary, lower values will add noise in the extracted sentences. A balance should be found, as in both cases it will be more difficult for the system to extract the characteristics of a style.

The proper threshold is thus selected through the evaluation metrics. Apart from the value of 0.6 in the direction of abstract to conclusion, no large variation is observed for classification accuracy. Now, a threshold value of 0.7 achieves the best results in both BLEU scores and perplexity in both directions. We thus use a threshold value of 0.7 for our next experiments.

TABLE II

STATISTICS ON UNPAIRED DATA CONSTRUCTION AND PREPARATION OF TRAINING

	Before filtering		After filtering		Experimental data		
	# of sentences	Avg. length	# of sentences	Avg. length	Training	Test	Dev.
Abstract	93,214	14.91	44,901	12.86	43,091	500	500
Conclusion	100,787	15.88	48,483	13.38	47,483	500	500

TABLE III

IMPACT OF THRESHOLD T ON UNPAIRED DATA CONSTRUCTION

Threshold T	Total # of sentences	Abstract $\rightarrow$ Conclusion			Conclusion $\rightarrow$ Abstract		
		Classification accuracy	BLEU	Perplexity	Classification accuracy	BLEU	Perplexity
0.5	173,898	64.13	24.77	20.38	63.90	23.09	26.54
0.6	129,589	68.45	25.11	18.73	67.12	26.78	23.66
0.7	93,384	76.66	**25.25**	**15.64**	73.33	**27.23**	**19.57**
0.8	61,203	75.03	19.55	27.12	75.84	18.92	39.67
0.9	31,602	79.07	11.09	133.65	75.12	14.32	192.29

TABLE IV

AUTOMATIC EVALUATION OF THE PROPOSED METHOD AND BASELINES

Model	Abstract $\rightarrow$ Conclusion			Conclusion $\rightarrow$ Abstract		
	Classification accuracy	BLEU	Perplexity	Classification accuracy	BLEU	Perplexity
CrossAlign [3]	68.21	19.82	27.12	63.53	14.31	39.66
DualRein [6]	52.13	12.84	34.42	58.90	9.49	55.09
StyleTrans [4]	75.32	**26.67**	21.48	**76.93**	22.43	29.82
Proposed method w/o data construction	63.66	22.59	18.54	61.35	21.26	24.58
Proposed method with data construction	**76.66**	25.25	**15.64**	73.33	**27.23**	**19.57**

TABLE V

MANUAL EVALUATION OF THE PROPOSED METHOD. TYPICAL AND RANDOMLY SELECTED EXAMPLES OF TRANSFERRED SENTENCES.

	Abstract $\rightarrow$ conclusion
Source (hand selected)	we propose a number of solutions to this problem .
CrossAlign	we proposed a number of solution to this scheme .
DualRein	we presents a number of solutions to the problem .
StyleTrans	we presented a number of measures to this problem .
Proposed method	we proposed a number of issues to the problem .
Source (hand selected)	finally we describe plans for future work .
CrossAlign	finally we describes <unk> for future work . (<unk> stands for unknown word.)
DualRein	finally we described results for future work .
StyleTrans	finally we described plans for future work .
Proposed method	finally we presented directions for future work .
Source (drawn at random)	empirical evaluation results demonstrate the utility of our <unk> system .
CrossAlign	empirical evaluation results demonstrate to perform better than <unk> propagation .
DualRein	empirical evaluation results demonstrate the number of our <unk> system .
StyleTrans	empirical evaluation results consistent the following of our <unk> system .
Proposed method	empirical evaluation results in the effectiveness of our <unk> system .
	Conclusion $\rightarrow$ abstract
Source (hand selected)	we plan to test this approach on other important nlp problems .
CrossAlign	we will to test this approach on other sentiment retrieval problems .
DualRein	we have to model this method on other subtle nlp problems.
StyleTrans	we describe to test this approach on a central nlp task .
Proposed method	we propose to this approach on several different nlp tasks .
Source (hand selected)	we presented a cross - lingual framework for finegrained opinion mining .
CrossAlign	we present a text - based conversations for , , ,
DualRein	we present a context - english tagging : such with with .
StyleTrans	we present a domain - specific synonym for chinese transliteration task .
Proposed method	we present a large - based approach for grammatical sentiment mining .
Source (drawn at random)	the selected words were all <unk> polysemous
CrossAlign	the selected features are all <unk> hierarchy
DualRein	the selected error are all <unk> polysemous
StyleTrans	the extracted words are all <unk> , .
Proposed method	the extracted sentences are all <unk> sentences .

2) Automatic Comparison with Other Methods: We use three unsupervised learning methods as baselines for comparison: CrossAlign [3], DualRein [6] and StyleTrans [4]. All of them have been applied to sentiment style transfer.

Table IV provides a comparison of our method with previous approaches. Our model offers the best compromise over all evaluation metrics. For variation in style, measured by binary style classification, only StyleTrans and our proposed method achieve more than 70% accuracy. For preservation of content, measured by BLEU, our proposed method is only beaten by StyleTrans in the abstract to conclusion direction. The gap between two directions in BLEU score is smaller than other approaches, it can stably preserving content due to our two generators. The perplexity of our method is lower by a large margin in both directions, which reflects better fluency in the transferred sentences.

As an ablation experiment, we measure the impact of not using the data construction method described in Section IV-B. This is shown in the second last row of Table IV. The performance is clearly worse than the method with data construction algorithm.

3) Manual Comparison with Other Methods: we inspected by hand some transferred sentences output by our method and other methods. Some examples selected by hand and drawn at random are shown in Table V. Confirming human knowledge (see Section I), we consistently observe changes in tenses (present for abstract and past for conclusion, shown in red in Table V) in all methods. However the changes in word usage (in blue in Table V) do not allow us to draw any clear conclusion.

VI. Conclusion

We addressed the problem of style transfer between abstracts and conclusions of scientific articles in the NLP field. The challenge was the absence of clearly defined styles. To address this challenge, we first built a data set of unpaired sentences from a collection of scientific articles by relying on sentence similarity measured by the cosine similarity between vector representations of sentences. We then designed a model which uses the `CycleGAN` architecture. In this model, the generator is implemented as a transformer.

Our results showed that our model actually captures style variation while ensuring fluency and preserving content. A manual inspection of the output showed that the main differences exhibited in the transferred sentences consist in tenses and word usage.

We intend to apply our method to other section types like Introduction, and to integrate it into a writing aid system so as to reduce the human effort in writing scientific articles

VII. Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP18K11446 entitled: “Natural language processing for academic writing in English”.

References

[1] T. Li and Y. Lepage, “Informative sections and relevant words for the generation of NLP article abstracts,” in AMJANLP, 2019, pp. 1281–1284. [Online]. Available: https://www.anlp.jp/proceedings/annual_meeting/2019/pdf_dir/P7-10.pdf

[2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, “Attention is all you need,” in NIPS, 2017, vol. 30, pp. 5998–6008. [Online]. Available: http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf

[3] T. Shen, T. Lei, R. Barzilay, and T. Jaakkola, “Style transfer from non-parallel text by cross-alignment,” in NIPS, 2017, pp. 6833–6844. [Online]. Available: https://papers.nips.cc/paper/7259-style-transfer-from-non-parallel-text-by-cross-alignment.pdf

[4] N. Dai, J. Liang, X. Qiu, and X. Huang, “Style transformer: Unpaired text style transfer without disentangled latent representation,” in ACL, 2019, pp. 5997–6007. [Online]. Available: https://www.aclweb.org/anthology/P19-1601

[5] S. Prabhumoye, Y. Tsvetkov, R. Salakhutdinov, and A. W. Black, “Style transfer through back-translation,” in ACL, 2018, pp. 866–876. [Online]. Available: https://www.aclweb.org/anthology/P18-1080

[6] F. Luo, P. Li, J. Zhou, P. Yang, B. Chang, Z. Sui, and X. Sun, “A dual reinforcement learning framework for unsupervised text style transfer,” in IJCAI 2019, 2019. [Online]. Available: https://www.ijcai.org/Proceedings/2019/0711.pdf

[7] J. Xu, X. Sun, Q. Zeng, X. Zhang, X. Ren, H. Wang, and W. Li, “Unpaired sentiment-to-sentiment translation: A cycled reinforcement learning approach,” in ACL, 2018, pp. 979–988. [Online]. Available: https://www.aclweb.org/anthology/P18-1090

[8] C. Wu, X. Ren, F. Luo, and X. Sun, “A hierarchical reinforced sequence operation method for unsupervised text style transfer,” in ACL, 2019, pp. 4873–4883. [Online]. Available: https://www.aclweb.org/anthology/P19-1482

[9] S. Rao and J. Tetreault, “Dear Sir or Madam, may I introduce the GYAFC dataset: Corpus, benchmarks and metrics for formality style transfer,” in NAACL-HLT, 2018, pp. 129–140. [Online]. Available: https://www.aclweb.org/anthology/N18-1012

[10] I. Czeresnia Etinger and A. W. Black, “Formality style transfer for noisy, user-generated conversations: Extracting labeled, parallel data from unlabeled corpora,” in W-NUT, 2019, pp. 11–16. [Online]. Available: https://www.aclweb.org/anthology/D19-5502

[11] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” in EMNLP-IJCNLP, 2019, pp. 3982–3992. [Online]. Available: https://www.aclweb.org/anthology/D19-1410

[12] R. He and J. McAuley, “Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering,” in ICWWW, 2016, pp. 507–517. [Online]. Available: https://doi.org/10.1145/2872427.2883037

[13] Z. Hu, Z. Yang, X. Liang, R. Salakhutdinov, and E. P. Xing, “Toward controlled generation of text,” in ICML, 2017, pp. 1587–1596. [Online]. Available: http://proceedings.mlr.press/v70/hu17e/hu17e.pdf

[14] K. Lin, D. Li, X. He, Z. Zhang, and M.-t. Sun, “Adversarial ranking for language generation,” in NIPS, 2017, vol. 30, pp. 3155–3165. [Online]. Available: http://papers.nips.cc/paper/6908-adversarial-ranking-for-language-generation.pdf

[15] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in ICCV, 2017.

[16] T. Qiao, J. Zhang, D. Xu, and D. Tao, “MirrorGAN: Learning text-to-image generation by redescription,” in CVPR, 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8953553

[17] G. Peng, Y. Yang, C. Tsai, and V. Soo, “Generate modern Chinese poems from news based on text style transfer using GAN,” in TAAI, 2019, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/8959907

[18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in NIPS, 2014, vol. 27, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[19] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of tricks for efficient text classification,” in EACL, 2017, pp. 427–431. [Online]. Available: https://www.aclweb.org/anthology/E17-2068

[20] K. Heafield, “KenLM: faster and smaller language model queries,” in WMT, 2011, pp. 187–197. [Online]. Available: https://www.aclweb.org/anthology/W11-2123/