MultiRL: A reinforcement learning framework for unparallel literary text multi-style transfer

Chenlu Wang
Industrial Engineering
Beijing Jiaotong University
Peking, China
20221333@bjtu.edu.cn

Jie Tang
Optics
Shanghai Synchrotron Radiation Facility, UCAS
Shanghai, China
jietang@mail.ustc.edu.cn

Jialu Liu
Educational Technology
University of Michigan
Ann Arbor, USA
lynneliu@umich.edu

Abstract

Unsupervised text multi-style transfer aims to solve the problem of achieving to transfer the underlying style of text while keeping its main content unchanged among multiple styles without parallel data. Based on the Dual Reinforcement Learning framework (Luo et al., 2019), in this paper we propose a framework named MultiRL that extends the framework between two targets to a dual learning framework among multiple targets. A style-free vector representation for contents is introduced as the intermediate step between transformations among different styles. We perform dual learning between multiple targets via the representation vector, where each target gets multiple rewards on both style accuracy and content preservation to be trained by reinforcement learning based on multiple mapping models. The final result is expected to achieve transfer between multiple styles without parallel data.

1 Introduction

The development of text style transfer has achieved unprecedented success for the past decade, with significant technological progress in Deep Learning (DL) (Krizhevsky et al., 2017) including Computer Vision (CV) and Natural Language Processing (NLP). Inspired by image style transfer in which neural networks are used to produce new images with contents from arbitrary photos and styles of well-known artworks (Gatys et al., 2016), text style transfer task (Hu et al., 2017) is to transfer a given piece of text into a certain style while preserving the original content as much as possible. In particular, the style-transfer generation of texts has a source sentence and the desired style as input and a ”transformed” sentence with similar content in the desired style as output. However, text style transfer can be more challenging than in the visual domain, due to insufficient information in sentences and the difficulty in defining a speech style. Besides, it is common that researchers cannot guarantee access to supervised data sets or parallel corpora across different styles. Therefore, the task of unsupervised text style with non-parallel data transfer draws more attention.

There is plenty of prior work done about style transfer between different stylistic attributes, for example, sentiments, formality, politeness, personal styles, and genres (Toshevska and Gievska, 2021). A mainstream way to complete the task is to separate style-independent content representation from the style-dependent representation of the text via adversarial training (Prabhumoye et al., 2018); however, a problem might arise that it is hard to keep the original content due to the difficulty to grab a style-independent content vector without parallel data (Lample et al., 2018; Luo et al., 2019). Also, another trend is that most of the past research concentrates on two-style transfer, especially from one characteristic to the opposite (e.g. negative to positive); however, there are so many more complicated circumstances in real-life applications. Code-switching in different domains targeted at different audiences is very common between diverse speech styles including humor, word choices, and formality in alignment with the situation, which leads to a necessity to consider transfer between multiple styles rather than binary tasks.

To solve these two problems, we present our model of unsupervised multi-style text transfer. Our aim is to achieve text transfer among different personal writing styles in English literature. Given a piece of narrative text, our model will transfer it into the distinguishable styles of famous authors such as Shakespeare and others. In order to realize this function, we combined the Multi-style Transfer model (Chen et al., 2021) with the Dual Reinforce Learning (DualRL) model (Luo et al., 2019) and proposed our own method. The proposed framework uses an intermediate vector V as the style-independent representation, which links separate styles together.

In summary, our contribution can be listed as follows:

- We propose a novel model to address the unsupervised multi-style text transfer task. Our framework takes advantage of DualRL and makes further improvements in introducing intermediate preresentation for easily transferring among different styles.
- We deal with the writing style of literary works, where there are less researches than traditional topics (e.g. sentiment transfer) in NLP.

2 Related work on Text Style Transfer

The first approaches towards text style transfer was formulated as text generation controlled by certain text attributes (Hu et al., 2017). Later it is recognized that in text style transfer tasks, an original input sentence is rewritten in a new style with the same explicit meaning – see e.g. Toshevska and Gievska (2021) for a recent review. The objective of applying style transfer models is to adjust, modify, and adapt the in which the input sentence with respect to particular style properties. Among all the tasks, the most widely discussed issue is the sentiment style transfer which adjusts the emotions conveyed in a sentence, for example, to change a negative comment into a positive one. To achieve the transfer, different types of models have been proposed.

2.1 Delete, Retrieve and Generate

One of the pioneering encoder-decoder models using simple reconstruction method is the "Delete, Retrieve and Generate" model (Li et al., 2018), which maximizes the probability distribution of the next word in the output sequence conditioned on the latent representation of the words in the input sentence. It was mainly designed for the sentiment modification task to remove the style markers with the encoder, retrieve a similar sentence in the corpus, and generate the output sentence by the decoder. In our model, this model is applied as a preposition module.

2.2 Dual Reinforcement Learning

Another important model which our proposed framework is based on is the Dual Reinforcement Learning model (DualRL) (Luo et al., 2019). The model was outstanding because it avoided the controversial process to separate style from content and generated good results.

By arguing that the two-step process to first separate style and content and then fuse the content with the target style had apparent limitations, DualRL raised a one-step mapping model. It is an advanced model for the two-style transfer task, where the researchers designed dual models of target-to-source and source-to-target mappings and then trained them alternatively with reinforcement learning. DualRL made no distinction between content and style, which solved the difficulty to distinguish them. It was trained on non-parallel data, and includes the style part and the content part in its reward function. This model outperformed all the other models on sentiment transfer tasks, but it was computationally expensive and hard to train. Meanwhile, the model was limited to the traditional two-style transfer tasks, which leaves room for further adjustment into the multi-style framework.

2.3 Style transfer for literary works

Compared to normal text style transfer, the task of targeting literary works, which we focus on in this article, can be even more challenging. To start with, personal writing styles of different authors are more difficult to capture, in comparison to sentiment transfer tasks where expressions indicating positive or negative tonality are often easier to be detected. Second, style transfer between authors separated in time by centuries could constitute an additional challenge due to more significant differences in language. For example, Shakespeare used Early Modern English in his writing, which was expressed in a distinguishable way from today's standard English (Jhamtani et al., 2017). Although literary text style transfer is an important subtask of text style transfer, it seems that only few research papers have aimed to contribute towards this goal (Jhamtani et al., 2017; Peng et al., 2019; Xu et al., 2012), which is one of the reasons why we took this direction.

2.4 Multi-style transfer

There are many scenarios when we need to transfer texts from one source style to multiple target styles. In this case, training separate models for each target style is not ideal considering training efficiency and storage overhead (Chen et al., 2021). Thus, some researchers have explored the possibilities of extending binary tasks in unsupervised text style transfer into multi-style applications.

A common method is to extract one content module and attach it to multiple style modules. Chen et al. (2021) proposed their model of a parameter-shared style-independent module, which is based on the Non-Autoregressive Transformer (NAT) for the convenience of training. Li et al. (2020) applied adversarial training with a latent decomposition scheme with a style code and a content code to achieve the one-to-many mapping with Generative Adversarial Networks (GAN) to minimize multiple losses. Unified Generative Adversarial Networks (UGAN) was also used for this task (Yu et al., 2020). However, the experiment results of the existing studies were far from satisfactory compared to others in conventional binary sentiment transfer tasks. For example, the BLEU score of Chen et al. (2021) was relatively low and Li et al. (2020) reported failed attempts in preserving the content when the inputs were lengthy sentences and transferring the style if the sentence contained novel symbols or complicated structure. One possible reason for their failure could be that in a complex sentence, it would be difficult to disentangle "style" and "content". We hence leverage the DualRL mechanism into our proposed model in order to overcome this obstacle.

3 Methodology

Motivated by DualRL, we propose our dual reinforcement learning model for unsupervised multi-style transfer in two forms: A simple form as the first step and a complete form as our desired model. We aim to illustrate through these models that introducing the intermediation corpus of another style not only inproves the quality of transferred sentences but also expands the ability of our model to be fit for text multi-style transfer tasks.

3.1 Problem formulation

Given three corpora X, Y, and Z with three different styles S_X, S_Y and S_Z, by inserting an input sentence with one of these styles, the transfer task is aimed to generate sentences expressed in the other two styles while preserving the content of the source sentence. The instance of X, Y, and Z is denoted as x, y and z, and the transferred sentences are noted as x', y' and z', or as x'', y'' and z''. These three corpora are non-parallel, so there are no gold pairs such as (x^i, y^j, z^k) which describe the same content with different styles. Therefore, our models will be trained in an unsupervised way.

3.2 Simple form: Matched pairs of DualRL

Aligned with the DualRL model for two styles (Chen et al., 2021), we present our simple model which also directly learns two one-step mappings between the two corpora of different styles. This model is mostly a point-to-point combination of DualRL models between different corpora. For example, if we expect to transfer from style S_X to style S_Y, we can first consider the direct transfer from X to Y. In this case, the forward model $f : X \to Y$ transfers the source sequence x with style S_x into a sequence y' with style S_y, while the backward model $g : Y \to X$ transfers the sequence y with style S_y into a sequence x' with style S_x. Meanwhile, considering that we may also use style Z as the intermediation of style transfer, the two models $g' : X \to Z$ and $g'' : Z \to Y$ are also applied. The proposed multi-style transfer model is presented in Fig 1.

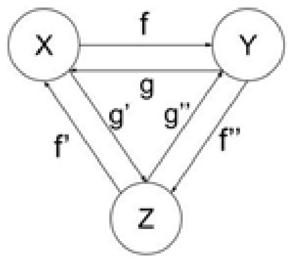

Figure 1: The simpler form of our model. Here X, Y and Z are non-parallel corpora, f and g are the same as those used in DualRL, where each pair of f and g (f' / g', f'' / g'') makes an intact DualRL model. We specify f/g to be Transformer (Vaswani et al., 2017) in our simple form.

3.2.1 Reward Design of DualRL

For the simple form of our model, there are two different parts of our reward function: the style reward for correctly transferring sentences into desired style R_s, and the reconstruction reward for correctly preserving the content of the original sentences in the style-transferred form R_c. Here we mostly use the same reward as that in DualRL, but we add one term $R_{c-inter}$ in R_c for the multi-style purpose.

In DualRL model, the backward model g is used statically when the forward model f is being updated; in our simpler model, all other than f is used statically when f is being updated. That way, one epoch consists of six separate training process, which separately train f, g, f', g', f'' and g'', rather than two process as those in DualRL. For the sake of clarity, here we assume we're training f; others rewards can be written by analogy.

For the style reward R_s, we use the probability of correctly classifying transferred y' as style S_Y using a discriminator pre-trained to distinguish among different possible styles S_X, S_Y and S_Z. It is explicitly expressed in Equation 1, where ϕ stands for the parameters of the f model.

$$R_s = P(S_Y/y'; \phi) \qquad (1)$$

The reconstruction reward R_c consists of two parts, and we note them as direct-transfer part $R_{c-direct}$ and intermediate-transfer part $R_{c-inter}$. If we use f to transfer original x into y', and then use g to transfer y' back into x', we can compare x and x' to get direct-transfer reconstruction reward by using cross entropy loss as in Equation 2, where $p(x)$ is a vector representation of sentence x and i denotes every component of this vector.

$$R_{c-direct} = -\sum_i p_i(x) \log(p_i(x')) \qquad (2)$$

On the other hand, if we use g' to transfer original x into z'', then use g'' to transfer z'' into y'', and lastly use g to transfer y'' into x'', we are able to further compare x and x'' to get intermediate-transfer reconstruction reward similarly by using cross entropy loss as in Equation 3.

$$R_{c-inter} = -\sum_i p_i(x) \log(p_i(x'')) \qquad (3)$$

Combining these rewards altogether with different weights a, b, c makes final reward R for the training process of f, as shown in Equation 4. Here we set a, b similar to that in DualRL, and adjust c so that introducing intermediate-transfer reconstruction reward not only preserves the main structure of DualRL but also achieves positive effect on the whole model.

$$R = aR_s + bR_{c-direct} + cR_{c-inter} \qquad (4)$$

3.2.2 Impossible Extension to Multiple Styles

In the simple form, it is relatively hard to generalize our model for fitting multiple styles other than three styles. For example, if we need a fourth style W in our model, we can only substitute, for example, Z and re-train new corresponding f', g', f'' and g'' to make it work among X, Y and W. However, this way it takes much time in re-training, and there is no way to transfer between Z and W. This problem only gets solved in the complete form of our model.

3.3 Complete form: Intermediate V for multi-style transfer

To achieve good model generalization for multi-style transfer task, we propose the complete form of our model with an intermediate vector V, as presented in Fig 2. In this model, instances of every corpora is first encoded into style-independent representation vectors V with corresponding encoder, and then decoded into desired sentences with transferred styles and preserved content. The intermediate vector V mitigates the pressure of building intact DualRL models between every pair of styles, thus accelerates the training process; to some extent, it also gives insights into style-independent representation of sentences.

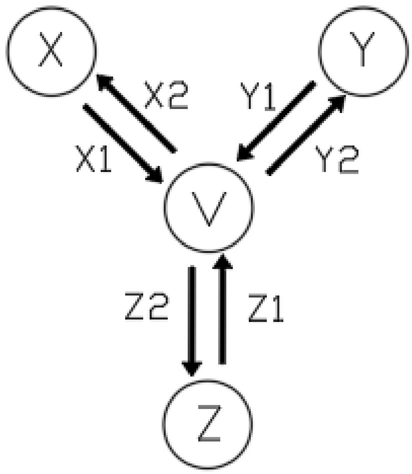

Figure 2: The complete form of our model. Here X, Y and Z are non-parallel corpora, X_1, Y_1 and Z_1 are encoders of Transformer with X_2, Y_2 and Z_2 being decoders of Transformer (Vaswani et al., 2017). These encoders and decoders are different for each corpus of different styles.

3.3.1 Our Proposed Reward Design

The reward function for the complete form consists of two parts, same as that in the simple form: the style reward R_s and the reconstruction reward R_c. Nevertheless, since there is an intermediate vector V, explicit forms of reward functions need to be adjusted.

In our complete form, we loop through all three corpora to accomplish one epoch, so without loss of generality, here we assume corpora X is now under consideration, which refers to the transferring from X to Y and from X to Z symmetrically. The reward for X is used to update the forward model X_1 and the backward model X_2. Rewards and updating processes for Y and Z can also be written by analogy.

The style reward R_s for the complete form consists of two parts: the style reward R_{s1} for correctly transferring from style S_X to style S_Y, and the style reward R_{s2} for correctly transferring from style S_X to style S_Z symmetrically. For the style reward R_{s1}, the instance x of corpora X is encoded by the forward model $X_1 : X \to V$, where intermediate vector v' is obtained; then by the backward model $Y_2 : V \to Y$,y' is obtained. As in the simple form, we use the probability of the pre-trained discriminator correctly classifying y' as style S_Y to be our style reward, as in Equation 5. Here ϕ denotes the parameters of all the X_i, Y_i and $Z_i (i = 1, 2)$

$$R_{s1} = P(S_Y/Y'; \phi) \tag{5}$$

Similarly, using the obtained v', by the backward model $Z_2 : V \to Z$,we can get z' and the corresponding style reward R_{s2} as in Equation 6:

$$R_{s2} = P(S_Z/Z'; \phi) \tag{6}$$

The content reward R_c for the complete form consists of three parts, noted as the direct-transfer part R_{c1}, the Y intermediate-transfer part R_{c2} and the Z intermediate-transfer part R_{c3}. These three parts are defined in analogy with the simple form using the cross entropy loss.

For the direct-transfer part R_{c1}, with v' obtained by the instance x and the forward model $X_1 : X \to V$, the backward model $X_2 : V \to X$ is applied to get x'. Here x' can be compared with x in the form of cross entropy to obtain the content reward R_{c1} as in Equation 7:

$$R_{c1} = -\sum_i p_i(x) \log(p_i(x')) \tag{7}$$

For the Y intermediate-transfer part R_{c2}, the y' (obtained by the vector v' and the backward model $Y_2 : V \to Y$) is passed back by the forward model $Y_1 : Y \to V$ to obtain v'', and then propagates through the backward model $X_2 : V \to X$ to get x''. This way, the content reward R_{c2} is constructed by the cross entropy loss between x'' and x, as in Equation 8:

$$R_{c2} = -\sum_i p_i(x) \log(p_i(x'')) \tag{8}$$

Similarly, for the Z intermediate-transfer part R_{c3}, the z' goes through the forward model $Z_1 : Z \to V$ to become v''', and then propagates through the backward model $X_2 : V \to X$, to be x'''. The content reward R_{c3} is the cross entropy loss between x''' and x as in Equation 9:

$$R_{c3} = -\sum_i p_i(x) \log(p_i(x''')) \tag{9}$$

Adding all the style reward terms and reconstruction reward terms, we can get the total reward R as in Equation 10. Here a, b and c are weights for different reward terms; due to symmetric consideration between Y and Z, the weights for R_{s1}, R_{s2} and the weights for R_{c2}, R_{c3} are the same respectively, and thus omitted.

$$R = aR_{c1} + b(R_{c2} + R_{c3}) + c(R_{s1} + R_{s2}) \tag{10}$$

3.3.2 Possible Extension to Multiple Styles

In our simple form, we are unable to generalize our model into multi-style scenarios. However, in this complete form, model generalization becomes possible in that there is intermediate V to help transfer among different styles. Whenever a new style, for example W, is added, training W_1 and W_2 that connect intermediate V and W alone is enough. To transfer towards new style W, a given sentence from any style (for example, y^* from Y) can be first encoded into intermediate V using corresponding Y_1, and then decoded into desired new style w^* using corresponding W_2. This way, less burden is loaded onto training process, and the model can be easily generalized to transfer among multiple styles given corresponding corpus. The transferring process between any pair of styles can be achieved without building intact DualRL models between them.

4 Experiments

4.1 Dataset

We choose Wikipedia XML data, modern translations of Shakespeare's plays and ASV version of Bibles as our datasets for training. We plan to expand to works by Chinese authors in the future. Here are some examples.

Wikipedia XML data:

1. The Khoirat spring is set amidst lush greenery behind the Holy Cross chapel.
2. Henderson was born in Brooklyn, New York.
3. No known queens from these dynasties.

Modern translations of Shakespeare's plays:

1. "Shall we to th' court?"
2. My lord, I have remembrances of yours That I have longèd long to redeliver.
3. My strength is all gone, as if in a dream.

ASV version of Bibles:

1. that can hurt you, if you be zealous of good?
2. For you are all the children of God by faith, in Christ Jesus.
3. We love, because he first loved us.

4.2 Training Details

Due to limited computational resources, our model is still under training and modification and we have not yet obtained satisfactory results.

4.3 Evaluation Metrics

Similarly to DualRL, we expect to conduct both automatic and human evaluation of our results.

Automatic Evaluation. Between X and Y, Y and Z, X and Z, we evaluated the accuracy of correctly classifying transferred sentences to the designated styles using our fine-tuned version of the pre-trained binary style discriminator TextCNN (Kim, 2014). The BLEU score (Papineni et al., 2002) is applied to evaluate the content preservation performance. Additionally, we introduced perplexity using a pre-trained language model (Zaremba et al., 2014) on PTB dataset (Marcus et al., 1993) to ensure the generated sentences are in natual expressions. In evaluating the overall performance, we expect to use the geometric mean of these three metrics.

Human Evaluation. We plan to hire several people with some language proficiency in English, and distribute the output sentences generated from different models including ours to them. Without knowing the system from which the generated text comes, They will be asked to score the generated text from 1 to 5 in terms of four criteria: the accuracy of the target style, the preservation of the original content, the fluency and subjective feelings as a human. Finally, following (Li et al., 2018), a transferred text is considered to be "successful" if it is rated 4 or 5 on all four criteria. In order to evaluate the quality of human evaluations,we will ask them to evaluate several times the same sample and look at the correlations between their evaluations (how much they agree), and consider improving the scoring rules based on this.

4.4 Expected results

Our model should enable transfer between multiple styles without parallel data. Compared to the most advanced systems, it will hopefully have a great advantage and lead, be more stable and perform better than DualRL in general.

References

Xi Chen, Song Zhang, Gehui Shen, Zhi-Hong Deng, and Unil Yun. 2021. Towards unsupervised text multi-style transfer with parameter-sharing scheme. *Neurocomputing*, 426:227–234.

Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2016. Image style transfer using convolutional neural networks. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423.

Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P. Xing. 2017. Toward controlled generation of text. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1587–1596. PMLR.

Harsh Jhamtani, Varun Gangal, Eduard Hovy, and Eric Nyberg. 2017. Shakespearizing modern language using copy-enriched sequence-to-sequence models. *arXiv preprint arXiv:1707.01161*.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90.

Guillaume Lample, Sandeep Subramanian, Eric Smith, Ludovic Denoyer, Marc'Aurelio Ranzato, and Y-Lan Boureau. 2018. Multiple-attribute text rewriting. In *International Conference on Learning Representations*.

Juncen Li, Robin Jia, He He, and Percy Liang. 2018. Delete, retrieve, generate: A simple approach to sentiment and style transfer. *arXiv preprint arXiv:1804.06437*.

Xiao Li, Guanyi Chen, Chenghua Lin, and Ruizhe Li. 2020. DGST: a dual-generator network for text style transfer. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7131–7136, Online. Association for Computational Linguistics.

Fuli Luo, Peng Li, Jie Zhou, Pengcheng Yang, Baobao Chang, Xu Sun, and Zhifang Sui. 2019. A dual reinforcement learning framework for unsupervised text style transfer. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 5116–5122. International Joint Conferences on Artificial Intelligence Organization.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Guan-Fu Peng, Yi-Shian Yang, Chen-Yu Tsai, and Von-Wun Soo. 2019. Generate modern chinese poems from news based on text style transfer using gan. In *2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, pages 1–6. IEEE.

Shrimai Prabhumoye, Yulia Tsvetkov, Ruslan Salakhutdinov, and Alan W Black. 2018. Style transfer through back-translation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 866–876, Melbourne, Australia. Association for Computational Linguistics.

Martina Toshevska and Sonja Gievska. 2021. A review of text style transfer using deep learning. *IEEE Transactions on Artificial Intelligence*, pages 1–1.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA. Curran Associates Inc.

Wei Xu, Alan Ritter, William B Dolan, Ralph Grishman, and Colin Cherry. 2012. Paraphrasing for style. In *Proceedings of COLING 2012*, pages 2899–2914.

Wei Yu, Tao Chang, Xiaoting Guo, Xiaodong Wang, Bo Liu, and Yang He. 2020. Ugan: unified generative adversarial networks for multidirectional text style transfer. *IEEE Access*, 8:55170–55180.

Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. 2014. Recurrent neural network regularization. *ArXiv*, abs/1409.2329.