Proceedings of the 4th International Conference on Computing Innovation and Applied Physics
DOI: 10.54254/2753-8818/86/2025.21002

A Study of Text Style Migration Based on Neural Machine Translation

Tao Yang[1,a,*]

[1]*College of Computer Science and Engineering, Changsha University, Changsha, 410022, China*
a. yt2797@ldy.edu.rs
**corresponding author*

Abstract: With the development of globalization, multilingual processing has emerged as a significant demand. Multilingual text-style migration has several real-world uses, such as improving intercultural communication, encouraging more inclusive multilingual education, and automating content production. These developments help to lower language barriers and promote global connectedness by making texts easier to read and accessible to a wide range of consumers. In the domain of natural language processing, the rapid growth of neural networks has led many scholars to devote themselves to the task of text-style migration using a neural machine translation system, facilitating the conversion of styles between different languages and enhancing the readability of texts. This paper first introduces the history of the development of neural machine translation and common neural machine translation modeling mechanisms. It then categorizes the text-style migration methods based on neural machine translation and specifies the applicable conditions, intrinsic mechanisms, and previous research results of each technique. This overview aims to shed light on the challenges and potential directions in the field of text-style migration, contributing to a deeper understanding of multilingual natural language processing. It is hoped that these revelations will stimulate more investigation and study in this developing field.

Keywords: NLP、machine translation、style transfer

1. Introduction

Style migration refers to converting text or images from one style to another, aiming to maintain the meaning of the content while changing its expression. This technique can enhance the diversity and adaptability of styles to meet the needs of different audiences. Gatys et al. were the first to propose the concept of style transfer, which pioneered the separation of the content and style of an image, and the application of an image's artistic style to the content of other images [1]. Based on this concept, the later proposed Text Style Transfer (TST), is to convert the expression or style of a text without changing the basic semantics of the text.

Traditional text-style migration methods, however, often lack a deep understanding of context. As a result, these methods may lose semantic consistency or cause the generated text to be less natural when dealing with complex text. Meanwhile, many traditional text style migration methods can only convert between specific styles, resulting in a single conversion result.

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/86/2025.21002

To make up for the shortcomings of the traditional text style migration, as deep learning has advanced, a text style migration technique based on Neural Machine Translation (NMT) has been proposed and gradually become mainstream. Prabhumoye et al. used a two-way translation network in NMT to transform the input sentences through different styles of translation models to realize text style migration and achieved relatively good translation results [2].

This paper first introduces neural machine translation and discusses the application scenarios of each model. After that, it will elaborate on the text style migration methods based on neural machine translation, categorize the different methods to introduce their mechanisms and application limitations. Finally, this paper will summarize the content of this paper and anticipate the future development trends in this field.

2. Introduction to neural machine translation

2.1. History of development

Machine translation is the process by which a computer converts a sentence in the source language into a sentence in the target language that is semantically equivalent. This is a significant area of study in the science of natural language processing. With the development of deep learning, researchers have started to use and then the vector c is decoded to generate a sequence of translation results $Y = (y1, y2, y3, \cdots, yn)$. The goal is to find the parameter θ that maximizes the conditeural networks to translate data, which is called NMT. NMT usually adopts the Encoder-Decoder Model (EDM), where a neural network is used to encode a source language sequence containing n words, to obtain a multidimensional vector c, and then a multidimensional vector c is obtained $X = (x1, x2, x3, \cdots, xn)$. First, a neural network is used to encode the source language sequence containing n words to obtain a multidimensional vector c, ional probability $P(Y|X)$ maximizes the conditional probability, which indicates the likelihood of the sentence being translated. Only a parallel corpus at the sentence level is needed for neural machine translation, and it is of high practical value to realize the translation process by simply adopting neural networks, which is convenient for training large-scale translation models. Currently, the most popular models in the field of machine translation are all neural machine translation models.

2.2. Neural Machine translation model

In neural machine translation, there are three mainstream models, namely Recurrent Neural Network based NMT (RNMT), which mostly uses recurrent neural networks as its foundation; Convolutional Sequence to Sequence Learning (ConvS2S) model, which primarily uses convolutional neural networks as its foundation; and Transformer model, which primarily relies on the Self-Attention process. These models have different advantages in their respective aspects.

The RNMT model has been a mainstream model of neural machine translation for a long time. Jean et al. proposed the RNNSearch model based on the encoder-decoder model, which introduces an attention mechanism so that when generating each word at the target end, the decoder can focus its attention on the relevant words at the source end and obtain useful information from them, thus obtaining better translation performance. The model introduces an attention mechanism so that when generating each target word, the decoder can focus its attention on the relevant words at the source and obtain useful information from them, thus obtaining better translation performance. The translation model can handle long-distance dependencies more effectively because of the attention mechanism, which also addresses the issue of information loss and forgetfulness during long-distance transmission in recurrent neural networks [3].

DOI: 10.54254/2753-8818/86/2025.21002

After the strong performance of neural machine translation models based on recurrent neural networks, researchers have begun to explore neural translation models based on convolutional neural networks, such as ConvS2S, which was first proposed by Facebook and is based on Convolutional Neural Networks (CNNs) and Attention mechanisms. CNNs can make full use of weight sharing and local correlation to greatly reduce the number of network parameters, thus making it easier to train deep neural networks on a large scale. CNN can take full advantage of weight sharing and local correlation to greatly reduce the number of network parameters, thus making it easier to realize large-scale deep neural network training. Traditional Natural Language Processing (NLP) tasks based on Recurrent Neural Network (RNN) cannot process data in parallel and must be computed one by one in time steps, which results in slower training speeds. One of the major advantages of CNN is that it can process data in parallel, making computation relatively more efficient.

Vaswani et al. proposed the Transformer model, which is completely based on the attention mechanism, which uses the self-attention mechanism to encode sequences, and its encoder and decoder are composed of an attention module and a forward neural network. An encoder and a decoder make up the Transformer model, and both have the same internal architecture, but with the addition of an extra cross-attention layer, which is the cross-attention module for the interaction between the encoder and the decoder. Due to its high level of parallelization, the Transformer model trains far more quickly than a recurrent neural network, and the effect of the translation has been greatly enhanced [4]. Recently, Transformer has become a mainstream model in neural machine translation research and has been extensively used in the field of NLP.

3. A neural machine translation-based approach to text style migration

The current mainstream classification methods classify NMT-based text style migration methods into three main categories, namely supervised learning methods, semi-supervised learning methods, and unsupervised learning methods. Supervised learning methods mainly use the seq-seq model to transform the translation problem. However, due to the absence of numerous parallel corpora, it is difficult to directly use the seq-seq model commonly used in machine translation for the text style migration task. Most of the studies aim at separating the potential representations of text attributes and text content through some learning mechanism from the perspective of unsupervised learning and then fusing the text content and the target attributes to realize the style migration.

3.1. Methods of supervised learning

The supervised learning approach aims to utilize the seq-seq model to train labeled data to predict the categories of the new input data when the parallel corpus is sufficient. When the translation model is being trained, the input to the encoder is a text sequence that needs to be converted to its style, and after converting the text sequence to context vectors, the decoder receives the context vectors and the labeling or embedding vectors of the target style, directs the generative process, and outputs the converted text to the target style.

In 2012, Xu et al. proposed a phrase-based machine translation system designed to convert modern English texts into Shakespearean play-style texts. This system utilizes phrase lists and conversion rules to identify and replace words and sentence patterns in modern English so that the output conforms to the linguistic features of Shakespeare's plays [5]. Later, given the quick development of neural network modeling, Jhamtani et al. used neural networks to pre-train word embeddings by constructing an external lexicon to map Shakespearean words to modern English words with limited parallel data, and used an end-to-end trainable neural model with pointers to transform texts from modern English to Shakespearean English, advancing the automated process of text style migration [6].

DOI: 10.54254/2753-8818/86/2025.21002

3.2. Methods for semi-supervised learning

The semi-supervised learning strategy aims to combine a small number of class-labeled samples with a large number of class-unlabeled samples in order to maximize the utilization of the unlabeled data and enhance the learning model's performance in the presence of a scarcity of massively parallel corpus, saving considerable translation time and cost.

Shang et al. proposed the Cross projection in latent space (CPLS) method, which is based on the seq-seq framework, and introduces a projection function to create a link of projection between various latent spaces and transform the intermediate vectors between two latent spaces using projection. Different constraints are designed to train the projection function for parallel and non-parallel corpora, respectively. The model contains two pairs of encoders and decoders, labeled A and B. For parallel corpus, a cross-projection strategy is designed. After encoding sentence an in sentence pair (a, b) as by encoder A, the projection function sends it to decoder B after projecting it into B's latent space. which will give the latent representation of b. For the non-parallel corpus, a cyclic projection strategy is designed to train the model by back-and-forth projection to train the projection functions to reconstruct the input [7].

In order to disambiguate Word Sense Disambiguation (WSD) in translation, Duarte et al. studied the use of neural network-generated word vector representations to capture the semantic differences of words in different contexts, which helps the model to more clearly differentiate the specific meanings of the words in the word sense disambiguation task. The method determines the specific meanings of words with multiple meanings represented in different contexts by embedding neural words, which effectively improves disambiguation accuracy [8].

3.3. Methods for unsupervised learning

Unsupervised learning approaches, which train models on non-parallel data by analyzing large amounts of unlabeled data to achieve textual style transformation, are the main methods explored for textual style transformation tasks today. One of the main obstacles in unsupervised translation tasks is the separation of content and style in a text. Depending on the form of the separated data, this class of methods can be subdivided into implicit and explicit methods.

3.3.1. Implicit methods

Implicit methods are those in which the model automatically captures the underlying features and stylistic representations of a sentence without explicitly identifying the words and phrases associated with the style. Implicit methods are usually based on models such as the generative adversarial network (GAN), variational auto-encoder (VAE), auto-encoder (AE), and so on. They also employ techniques like disentanglement, reinforcement learning, adversarial training, back-translation, and so forth to uncover the text's implicit representation.

The de-entanglement strategy mainly involves mapping the text to the hidden space to get the latent representations through an encoder, thus separating the content from the attributes and performing attribute migration. Hu et al.'s model builds on VAE, aiming to generate sentences that match the target attributes by learning the latent representations of style and content de-entanglement. The model does not enforce the consistency of the transformed content, and thus the degree of content preservation is poor [9]. Shen et al. proposed a novel approach, an unsupervised learning model based on RNN, to align the generator's hidden states at each time step during reconstruction and transformation as inputs to the discriminator, to share the same latent space of textual content hidden variables of different styles by utilizing the loss of the KL dispersion and to style transformation of text through adversarial training of the generator and the discriminator [10]. The VAE-based model proposed by Mueller et al. utilizes hidden representations to generate sentences that satisfy desired

DOI: 10.54254/2753-8818/86/2025.21002

styles, and these style attributes are measured by the corresponding classifier metrics but do not enforce content consistency [11]. Several research works have de-entangled the representation of text style and content in potential space by constructing cyclic reconstruction loss and adversarial loss for both encoder and generator. Logeswaran et al. pointed out that reconstruction prefers to replicate the source sentence and cyclic reconstruction may lead to an increased discrepancy in content, and proposed to construct a method of inserting a reconstruction loss between the two losses to ensure that the model generates sentences that are consistent with the content method to ensure that the model generates content-compatible sentences [12].

A machine learning problem known as reinforcement learning method is one in which an intelligent system continuously interacts with its surroundings to determine the best behavioral tactics and maximize the rewards accumulated over time. In the field of textual style migration, different reward function mechanisms can be designed to facilitate models to better learn textual content and stylistic representations. Luo et al. suggested a framework for dual reinforcement learning by considering learning two tasks: learning source-to-target and target-to-source mappings; and creating two rewards based on this dual structure, which represent, respectively, stylistic accuracy and content retention [13]. Reinforcement learning can be used to train these two one-step mapping models in this manner without the need for parallel data.

In neural machine translation, back translation is a frequently employed technique to make better use of monolingual corpora to aid in the training of translation models. It is based on the principle of translating the text from the source language into a translation and then translating the translation back to the source language. Although the translated text may contain unknown words or grammatical errors, since the source text is a high-quality monolingual corpus, the training process can be interpreted as a de-noising training method. Prabhumoye et al. introduced the idea of back-translation into the text style migration task [2]. They first use a machine translation model to translate the source text into another language, which is encoded to get), where denotes a sentence in language f, and denotes the parameters of the encoder, and the corresponding decoder can be used to generate sentences of the corresponding style. However, back-translation mainly improves the translation of inputs with similar styles. For natural inputs, back-translation can only bring slight improvements and sometimes even cause unfavorable effects. Therefore, Wei et al. proposed Text Style Migration Backtranslation (TST BT), which points out the difference between Nature, data generated by native speakers, MT, data generated by human translators, and HT, data generated by machine translators [14]. Lample et al. introduced a pooling operator in the latent space to provide greater control over the balance between changing the style and preserving the content, and to provide a more efficient way of text rewriting in complex scenarios. This provides new ideas for text rewriting in complex scenarios [15].

3.3.2. Explicit methods

Display methods consider that the stylistic features of a sentence are usually embodied in specific phrases and aim to directly identify and modify specific attribute words in a sentence for textual style migration. Currently, word-frequency-based deletion strategies, attention mechanism-based deletion strategies, and deletion strategies based on the combination of word-frequency and attention mechanisms are mainly used.

Word Frequency-Based Deletion Strategies The core idea of the strategy is to analyze word frequency to identify high-frequency words associated with a particular style or sentiment attribute and remove them from the sentence. Li et al. were the first to propose the delete - retrieve - enerate (DRG) framework for sentiment and style migration. The method's central concept is to convert a sentence's style or sentiment attribute to a target style or sentiment attribute in three steps while

DOI: 10.54254/2753-8818/86/2025.21002

keeping the sentence's primary idea intact. The DRG method simplifies the complex task of style migration and achieves efficient style conversion through a modular approach [16].

Introducing an attention mechanism in neural machine translation permits the model to generate Every word in the intended phrase while dynamically focusing on various elements of the source sentence. This advantage over the traditional seq2seq model, which has difficulty in compressing the complete input series into a fixed-length vector when dealing with complex sentences. Bahdanau et al. added a layer of attention in the middle of the machine translation model, which simultaneously translates and aligns phrases from the language of origin to the language of destination [17]. Recently, the attention mechanism has also appeared in text style migration for attribute word deletion. Xu et al. proposed a recurrent reinforcement learning method, which is trained on the data through the collaboration between a neutral module and an affective module. Emotion-to-emotion translation is accomplished by the emotionalization module, which adds emotions to the neutralized semantic material after the neutral module has explicitly filtered out emotionally charged terms. Meanwhile, a gradient strategy is adopted to reward the output of the neutral module based on the feedback from the emotionalization module [18].

Wu et al. take a two-step approach, masking and filling, to modify sentence sentiment attributes through a deep two-way masking language model. In the masking step, The substance and style are distinct by masking the position of sentiment markers. In the filling step, the mask language model (MLM) is modified to an attribute-conditional MLM model, which fills the masked positions by word or phrase prediction based on target mood and context to achieve the goal of text style migration [19].

4. Conclusions

This paper first outlines the history of neural machine translation development and its model architecture, then provides a categorized review of neural machine translation-based text style migration methods, and elaborates on the core mechanisms of each method in detail. However, there are still two major challenges in this field, the first is the scarcity of high-quality parallel corpus, which seriously restricts the training effect and performance improvement of the model; and the second is the lack of a unified and efficient evaluation system to objectively measure the effectiveness of the style migration, as people's definitions of styles are often highly subjective. Looking ahead, researchers in this field expect to build more diverse style migration datasets to better support the training and learning of the model and expect to realize text-style migration in more languages to promote the diversified development of this field.

References

[1] *L.-A. Gatys, A.-S. Ecker, M. Bethge, Image style transfer using convolutional neural networks, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2016), 2414-2423*

[2] *S. Prabhumoye, Y. Tsvetkov, R. Salakhutdinov, et al. Style transfer through back-translation. arXiv preprint arXiv:1804.09000 (2018)*

[3] *S. Jean, K. Cho, R. Memisevic, et al. On using very large target vocabulary for neural machine translation. arXiv preprint arXiv:1412.2007 (2014)*

[4] *Vaswani, N. Shazeer, N. Parmar, et al., Attention Is All You Need. In Advances in Neural Information Processing Systems. arXiv preprint arXiv:1706.03762, 10: S0140525X16001837 (2014)*

[5] *W. Xu, A. Ritter, W.-B. Dolan, et al. Paraphrasing for style, in Proceedings of COLING 2012, (2012), 2899-2914*

[6] *H. Jhamtani, V. Gangal, E. Hovy, et al. Shakespearizing modern language using copy-enriched sequence-to-sequence models. arXiv preprint arXiv:1707.01161 (2017)*

[7] *M. Shang, P. Li, Z. Fu, et al. Semi-supervised text style transfer: cross projection in latent space. arXiv preprint arXiv:1909.11493 (2019)*

[8] *J.-M. Duarte, S. Sousa, E. Milios, et al., Deep analysis of word sense disambiguation via semi-supervised learning and neural word representations. Inf. Sci. 570, 278-297 (2021)*

DOI: 10.54254/2753-8818/86/2025.21002

[9] *Z. Hu, Z. Yang, X. Liang, et al., Toward controlled generation of text, in Proceedings of the International Conference on Machine Learning, PMLR, (2017), 1587-1596*
[10] *T. Shen, T. Lei, R. Barzilay, et al., Style transfer from non-parallel text by cross-alignment. Adv. Neural Inf. Process. Syst. 30 (2017)*
[11] *J. Mueller, D. Gifford, T. Jaakkola, Sequence to better sequence: continuous revision of combinatorial structures, in Proceedings of the International Conference on Machine Learning, PMLR, (2017), 2536-2544*
[12] *L. Logeswaran, H. Lee, S. Bengio, Content preserving text generation with attribute controls. Adv. Neural Inf. Process. Syst. 31 (2018)*
[13] *F. Luo, P. Li, J. Zhou, et al. A dual reinforcement learning framework for unsupervised text style transfer. arXiv preprint arXiv:1905.10060 (2019)*
[14] *D. Wei, Z. Wu, H. Shang, et al. Text style transfer back-translation. arXiv preprint arXiv:2306.01318 (2023)*
[15] *G. Lample, N. Zeghidour, N. Usunier, et al., Fader networks: manipulating images by sliding attributes. Adv. Neural Inf. Process. Syst. 30 (2017)*
[16] *J. Li, R. Jia, H. He, et al. Delete, retrieve, generate: a simple approach to sentiment and style transfer. arXiv preprint arXiv:1804.06437 (2018)*
[17] *D. Bahdanau, Neural machine translation by joint learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)*
[18] *J. Xu, X. Sun, Q. Zeng, et al. Unpaired sentiment-to-sentiment translation: a cycled reinforcement learning approach. arXiv preprint arXiv:1805.05181 (2018).*
[19] *X. Wu, T. Zhang, L. Zang, et al. Mask and Infill: applying Masked Language Model to Sentiment Transfer. arXiv preprint arXiv:1908.08039 (2019)*